An Expectation-Maximization Algorithm for Bayesian Operational Modal Analysis with Multiple (Possibly Close) Modes

Binbin Li [a,b*], Siu-Kui Au [c]

[a]ZJU-UIUC Institute, Zhejiang University, Haining, 314400, China. Email: binbinli@intl.zju.edu.cn
[b]College of Civil Engineering and Architecture, Zhejiang University
Hangzhou, 310058, China. Email: bbl@zju.edu.cn
[c]School of Civil and Environmental Engineering, Nanyang Technological University, Email: ivanau@ntu.edu.sg
Formerly with
Institute for Risk & Uncertainty and Center for Engineering Dynamics, University of Liverpool,
Liverpool, L69 3GH, UK. Email: siukuiau@liverpool.ac.uk

Abstract: The Bayesian FFT method has gained attention in operational modal analysis of civil engineering structures. Not only the most probable value (MPV) of modal parameters can be computed efficiently, but also the identification uncertainty can be rigorously quantified in terms of posterior covariance matrix. A recently developed fast algorithm for general multiple (possibly close) modes was found to work well in most cases, but convergence could be slow or even fail in challenging situations. The algorithm is also tedious to computer-code. Aiming at resolving these issues, an expectation-maximization (EM) algorithm is developed by viewing the modal response as a latent variable. The parameter-expansion EM and the parabolic-extrapolation EM are further adopted, allowing mode shape norm constraints to be incorporated and accelerating convergence, respectively. A robust implementation is provided based on the QR and Cholesky decompositions, so that the computation can be done efficiently and reliably. Empirical studies verify the performance of the proposed EM algorithm. It offers a more efficient and robust (in terms of convergence) alternative that can be especially useful when the existing algorithm has difficulty to converge. In addition, it opens a way to compute the MPV in the Bayesian FFT method for other unexplored cases, e.g., multi-mode multi-setup problem.

Key words: Operational modal analysis; Bayesian inference; Expectation Maximization; Closely-spaced modes

1 Introduction

Operational modal analysis (OMA) [1–3] aims at identifying structural modal parameters (e.g., modal frequencies, damping ratios and mode shapes) by using only the measured structural response. The identified modal parameters play an important role in structural design verification and retrofits [4,5], vibration control [6,7] and health monitoring [8,9]. Since OMA does not require specific knowledge of the input force, it has gained popularity in the dynamic testing of civil engineering structures, where artificial excitation is costly or in many cases impractical.

In OMA, the input force is unknown and usually modeled as a wideband stochastic process. As a result, the identification process is more sophisticated compared to the input-output

* Corresponding author

identification, and proper means must be devised to extract modal information from the stochastic response data in the absence of loading information. During the past decades, a large variety of methods have been developed in literature. Early methods such as random decrement [10] and peak-picking [11] are simple to use but they can only provide a rough estimate of the modal parameters. During the 1990s, two important techniques became available: the stochastic subspace identification [12] and the frequency-domain decomposition [13]. They are able to estimate modal parameters with good accuracy in most conditions. Various algorithms [14–16] have been recently proposed to quantify the identification uncertainty in a frequentist sense by means of perturbation. For a more thorough overview of OMA methods, the reader is referred to Refs. [1–3]

Uncertainty quantification is central to Bayesian OMA methods [17], which addresses simultaneously estimation and uncertainty quantification via Bayes' Theorem for given measured data and modeling assumptions. It views modal identification as an inference problem, where identification result is encapsulated in the posterior probability density function (PDF) of modal parameters for given data and modeling assumptions. Based on different models and assumptions, several Bayesian OMA methods have been proposed, e.g., [18–20] in the time domain and [21–27] in the frequency domain. Balancing robustness in modeling assumptions and computational efficiency, the Bayesian FFT (Fast Fourier Transform) method [22–26] provides a feasible strategy for Bayesian analysis of modal parameters. The fast algorithm developed in [25] is applicable for general multiple (possibly close) modes. Compared to the algorithm for well-separated modes [23], the theory behind and the resulting programing effort were significantly more involved. It was based on heuristic grouping of parameters and iterative updating until convergence. It was found to work well in most applications, but the convergence can be an issue in challenging situations, e.g., when there are three or more modes in the selected frequency band. This paper presents an expectation-maximization (EM) algorithm, which does not involve heuristic grouping and offers a more robust (in terms of convergence) alternative for efficiently computing the most probable value (MPV) for general multiple (possibly close) modes.

The EM algorithm [28] is a popular tool for the maximum likelihood or maximum a posteriori estimation of statistical models in science and engineering. An up-to-date treatment of the theory and applications of this algorithm can be found in [29]. In the context of OMA, it has been applied in [30,31] to identify the state-space model. The EM algorithm is simple to implement and converges monotonically in terms of the loglikelihood or log-posterior of the latent-variable model [32] under mild conditions [28,33]. However, the EM algorithm is known to converge slowly in some situations. This important aspect has received much attention recently and many algorithms have been proposed to accelerate convergence while preserving algorithmic simplicity. Two such acceleration methods are incorporated in this work, namely, the parameter-expansion EM [34] and the parabolic extrapolation EM [35].

This paper is outlined as follow. The Bayesian FFT method is first reviewed, including the modeling assumptions and the Bayesian formulation. The EM algorithm and its variants are then briefly introduced and tailored to the Bayesian FFT method for fast computation. The closed-form update of the E step and M step are derived, and a robust implementation is provided. Finally, empirical studies based on synthetic, laboratory and field test data are presented to illustrate the performance of the proposed algorithm.

2 Bayesian FFT method

This section reviews the Bayesian formulation for OMA using the scaled FFT of ambient vibration data. This involves stating the modeling assumptions, the parameters to be identified and deriving the likelihood function, which is practically the posterior PDF of parameters.

Let the time history data measured at n degrees of freedom (DoFs) of a structure under ambient vibration be $\{\hat{\boldsymbol{y}}_j \in \mathbb{R}^n : j = 0,1,\dots,N-1\}$ and abbreviated as $\{\hat{\boldsymbol{y}}_j\}$, where N is the number of samples per data channel and n is the number of data channels. Modeled as a stationary stochastic process, its scaled FFT is defined as

$$\hat{\boldsymbol{\mathcal{F}}}_k = \sqrt{\Delta t/N} \sum_{j=0}^{N-1} \hat{\boldsymbol{y}}_j e^{-2\pi \mathbf{i} jk/N} \tag{1}$$

where Δt (sec) is the sampling interval and $\mathbf{i}$ is the imaginary unit. For $k \le N_q$, $\hat{\boldsymbol{\mathcal{F}}}_k$ corresponds to the frequency $\mathrm{f}_k = k/N\Delta t$ (Hz), where $N_q = \mathrm{int}[N/2]$ ($\mathrm{int}[\cdot]$ denotes the integer part) is the index at the Nyquist frequency. The FFT in Eqn. (1) is scaled by the factor $\sqrt{\Delta t/N}$ so that the expectation of $\hat{\boldsymbol{\mathcal{F}}}_k \hat{\boldsymbol{\mathcal{F}}}_k^*$ ('*' denotes conjugate transpose) is equal to the two-sided power spectral density (PSD) matrix of the data process.

In practice, only the FFTs on a selected frequency band with N_f FFT points, denoted by $\{\hat{\boldsymbol{\mathcal{F}}}_k\}$, containing the mode(s) of interest are used for identification. This trades off between the information used for identification (the wider the better) and modeling error risk (the narrower the better) [36]. Within the selected band, it is assumed that

$$\hat{\boldsymbol{\mathcal{F}}}_k = \boldsymbol{\mathcal{F}}_k + \boldsymbol{\varepsilon}_k \tag{2}$$

where $\boldsymbol{\mathcal{F}}_k$ and $\boldsymbol{\varepsilon}_k$ denote respectively the scaled FFT of the theoretical structural dynamic response and the prediction error (e.g., data noise). Suppose the selected frequency band is dominated by m vibration modes, referred as mode $1,2,\dots,m$, then $\boldsymbol{\mathcal{F}}_k = \boldsymbol{\Phi}\boldsymbol{\eta}_k$ where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_m] \in \mathbb{R}^{n\times m}$ is the partial mode shape matrix confined to measured DoFs, and $\boldsymbol{\eta}_k \in \mathbb{C}^{m\times 1}$ is the scaled FFT of modal response at frequency f_k. Let $\boldsymbol{p}_k \in \mathbb{C}^{m\times 1}$ be the scaled FFT of the modal excitation at frequency f_k. Then $\boldsymbol{\eta}_k = \boldsymbol{h}_k \boldsymbol{p}_k$ and

$$\hat{\boldsymbol{\mathcal{F}}}_k = \boldsymbol{\Phi}\boldsymbol{h}_k\boldsymbol{p}_k + \boldsymbol{\varepsilon}_k \tag{3}$$

where $\boldsymbol{h}_k = \mathrm{diag}(h_{1k}, h_{2k}, \dots, h_{mk})$ is a diagonal matrix with frequency response functions (FRFs) as its diagonal elements. The FRF, for mode i, is given by

$$h_{ik} = \frac{(2\pi \mathbf{i}\mathrm{f}_k)^{-q}}{(1-\beta_{ik}^2) - \mathbf{i}(2\zeta_i\beta_{ik})} \quad \beta_{ik} = \frac{f_i}{\mathrm{f}_k} \quad q = \begin{cases} 0, & \text{acceleration data} \\ 1, & \text{velocity data} \\ 2, & \text{displacement data} \end{cases} \tag{4}$$

where f_i (Hz) and ζ_i are respectively the ith natural frequency and damping ratio. In deriving these equations, we assume a linear time-invariant dynamic system with the classical damping. As it is widely recognized, the damping mechanism is generally complex and hard to model exactly. The classical damping assumption provides a mathematically simple damping model, that has been used conventionally for analysis and design of civil engineering structures.

In OMA, the physical excitation is not measured. The modal excitation $\boldsymbol{p}_k$ is therefore unknown but modeled statistically. Assuming zero-mean stationary modal excitation and long data ($N_f \gg 1$), $\boldsymbol{p}_k$ has a (circularly symmetric) complex Gaussian distribution and independent at different frequencies [37]. Its covariance (i.e., PSD of the process) is assumed to be a constant matrix $\boldsymbol{S} \in \mathbb{C}^{m\times m}$ in the selected band, i.e., $\boldsymbol{p}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{S})$. The prediction errors $\boldsymbol{\varepsilon}_k$ at different channels are also assumed to be complex Gaussian distributed with zero mean and covariance matrix $S_e\boldsymbol{I}_n$ ($\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ denotes the identity matrix) within the selected band, which can be justified based on the principle of maximum entropy [38]. Further assuming statistical independence between the modal excitation and prediction errors yields a jointly independent complex Gaussian distribution for $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ with zero mean and each with a covariance matrix (or theoretical PSD of data) given by

$$\boldsymbol{E}_k = \boldsymbol{\Phi}\boldsymbol{H}_k\boldsymbol{\Phi}^{\mathrm{T}} + S_e\boldsymbol{I}_n \tag{5}$$

where $\boldsymbol{H}_k = \boldsymbol{h}_k\boldsymbol{S}\boldsymbol{h}_k^*$ and '$\cdot^{\mathrm{T}}$' denotes real transpose. The covariance matrix $\boldsymbol{E}_k$ is central to OMA because it is the only channel through which the likelihood function is affected by the unknown parameters $\boldsymbol{\theta} = \{\boldsymbol{f}, \boldsymbol{\zeta}, \boldsymbol{\Phi}, \boldsymbol{S}, S_e\}$ where $\boldsymbol{f} = [f_1, f_2, \dots, f_m]^{\mathrm{T}} \in \mathbb{R}^{m\times 1}$ and $\boldsymbol{\zeta} = [\zeta_1, \zeta_2, \dots, \zeta_m]^{\mathrm{T}} \in \mathbb{R}^{m\times 1}$.

The complex Gaussian distribution for $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ is supported by the theoretical result that the scaled FFTs of a stationary process are asymptotically independent complex Gaussian distributed for long data [37]. This result is robust by virtue of the Central Limit Theorem and the fact that the FFT, response, modal force, physical force are successively linear combinations of the next [3]. The above frequency-domain model only makes use of the FFT information in the selected band, which significantly simplifies the identification model. The PSD of the modal excitation and prediction error need only be flat within the selected band, relaxing the conventional white noise assumption and making the method more robust than time-domain methods. Other bands with irrelevant information or which are difficult to model are legitimately ignored, therefore avoiding modeling error. This does not require any signal pre-processing such as filtering or averaging.

Using the above result for the likelihood function and adopting a uniform prior distribution, the posterior distribution of $\boldsymbol{\theta}$ is given by

$$p\left(\boldsymbol{\theta}\middle|\{\widehat{\boldsymbol{\mathcal{F}}}_k\}\right) \propto p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k\}\middle|\boldsymbol{\theta}\right) = \frac{\pi^{-nN_f}}{\prod_k|\boldsymbol{E}_k|}\exp\left[-\sum_k \widehat{\boldsymbol{\mathcal{F}}}_k^*\boldsymbol{E}_k^{-1}\widehat{\boldsymbol{\mathcal{F}}}_k\right] \tag{6}$$

where $|\cdot|$ denotes the matrix determinant and $\boldsymbol{E}_k$ is given in Eqn. (5).

Note that the problem is not identifiable because the external excitation is not measured and thus the mode shape matrix $\boldsymbol{\Phi}$ can be arbitrarily normalized. A conventional treatment is to introduce the norm constraints $\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{\phi}_i = 1$ for $i = 1,2,\dots,m$, which makes the problem locally identifiable as the sign of $\boldsymbol{\phi}_i$ can be positive or negative but is otherwise immaterial. A second order Taylor approximation of the log posterior PDF at the local maximum (referred as MPV hereafter) leads to a Gaussian approximation of the posterior PDF:

$$p\left(\boldsymbol{\theta}\middle|\{\widehat{\boldsymbol{\mathcal{F}}}_k\}\right) \approx (2\pi)^{-n_\theta/2}|\widehat{\boldsymbol{C}}|^{-1/2}\exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)^{\mathrm{T}}\widehat{\boldsymbol{C}}^{-1}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)\right] \tag{7}$$

where $\widehat{\boldsymbol{\theta}}$ is the MPV, $\widehat{\boldsymbol{C}}$ is the covariance matrix, equal to the negative inverse of the Hessian of the log posterior PDF at the MPV, $n_\theta = (m+1)^2 + mn$ is the number of parameters in $\boldsymbol{\theta}$. Both $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{C}}$ depends on the FFT data $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ and can be calculated when they are given.

The computation of the MPV $\widehat{\boldsymbol{\theta}}$ and the covariance matrix $\widehat{\boldsymbol{C}}$ is not a trivial task. Brute-force numerical optimization is computationally prohibitive, primarily because the number of modal parameters n_θ can be large. In addition, the computation is vulnerable to ill-conditioning problems because the matrix $\boldsymbol{E}_k$ in Eqn. (5) is almost rank-deficient when the signal-to-noise ratio is high (i.e., small S_e) and the number of modes m is less than that of the measured DoFs n. In view of the computational problems, efficient algorithms have been proposed in [23,24] for well-separated modes and [25,26] for general multiple (possibly close) modes, which allow the MPV and posterior covariance matrix to be computed typically in a matter of seconds. Since these algorithms directly optimize the posterior distribution of $\boldsymbol{\theta}$ shown in Eqn. (6), they are referred as the "direct method" hereafter. They have been found to work well in most applications, but convergence can sometimes become an issue in challenging situations, e.g., when there are three or more modes in the selected frequency band. In view of this, an EM algorithm is developed in this work to compute the MPV. With the premise of its monotonic convergence, it can provide an alternative means when the existing method fails to converge. Its simplicity can also significantly relieve computer programming effort.

3 Expectation maximization and its variants

Before applying to develop the proposed algorithm, we first give a brief introduction to the EM algorithm and its variants that will be used in the paper. The EM algorithm [28] is widely used to infer the statistical model that can be formulated as a latent variable model [32]. In the context of Bayesian FFT method, the measurement $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ is the observed variable and the modal response $\{\boldsymbol{\eta}_k\}$ can be regarded as the latent variable. Suppose that we have derived the conditional joint distribution of $\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}$ given $\boldsymbol{\theta}$, i.e., $p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\middle|\boldsymbol{\theta}\right)$. The EM algorithm computes the MPV $\widehat{\boldsymbol{\theta}}$ by iteratively proceeding, given a starting value $\boldsymbol{\theta}^{(0)}$, between the expectation (E) step and the maximization (M) step:

E step. Compute the expected complete-data log-likelihood

$$Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = \mathrm{E}_{\{\boldsymbol{\eta}_k\}|\{\widehat{\boldsymbol{\mathcal{F}}}_k\},\boldsymbol{\theta}^{(t)}}\left[L\left(\boldsymbol{\theta}\middle|\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\right)\right] \tag{8}$$

where $L\left(\boldsymbol{\theta}\middle|\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\right) = \log p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\middle|\boldsymbol{\theta}\right)$ and '$\mathrm{E}_{\{\boldsymbol{\eta}_k\}|\{\widehat{\boldsymbol{\mathcal{F}}}_k\},\boldsymbol{\theta}^{(t)}}[\cdot]$' is the expectation operation with respect to (w.r.t.) the conditional distribution of $\{\boldsymbol{\eta}_k\}$ given $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ and $\boldsymbol{\theta}^{(t)}$; and

M step. Maximize $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right)$ to obtain

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) \tag{9}$$

The preceding steps are repeatedly implemented until convergence is achieved. Wu [33] provides conditions under which $\{\boldsymbol{\theta}^{(t)}\}$ converges to the target optimum $\widehat{\boldsymbol{\theta}}$, which can be either a saddle point or a local maximum. The EM algorithm takes advantage of the model structure to decouple the optimization problem into more manageable pieces. This divide-and-conquer strategy gives a conceptual clarity and simplicity of the algorithm. It is particularly useful when the likelihood belongs to the exponential family [39]: the E step becomes the sum of expectations of sufficient statistics, and the M step involves maximizing a linear function. In such a case, it is usually possible to derive a closed-form update for each step.

While simple to implement and stable in its convergence, depending on the problem nature and the choice of latent variables, the EM algorithm can converge slowly. Many variants of the ordinary EM algorithm have been proposed in order to overcome shortcomings that are sometimes seen in implementations of the ordinary method. In this paper, we adopt two acceleration methods: the parameter-expansion EM (PX-EM) [34] and the parabolic-extrapolation EM (P-EM) [35]. The PX-EM is applied here to deal with the norm constraints on mode shapes, which forbids the closed-form update of mode shapes in the ordinary EM. By introducing an auxiliary variable, the efficient update of mode shapes becomes possible during PX-EM iterations. The P-EM is then employed to further accelerate the PX-EM algorithm.

The PX-EM algorithm is essentially an EM algorithm but it performs inference on a larger full model. This model is obtained by introducing extra parameters into the complete-data model $p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\middle|\boldsymbol{\theta}\right)$ while preserving the observed-data sampling model $p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k\}\middle|\boldsymbol{\theta}\right)$. Suppose that the EM complete-data model can be embedded in a larger model $p_*\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\middle|\boldsymbol{\theta}_*, \boldsymbol{\gamma}\right)$ with the expanded parameter set $\{\boldsymbol{\theta}_*, \boldsymbol{\gamma}\}$, where $\boldsymbol{\theta}_*$ plays the same role as $\boldsymbol{\theta}$ in $p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\middle|\boldsymbol{\theta}\right)$ and $\boldsymbol{\gamma}$ is the auxiliary parameter whose value is fix at $\boldsymbol{\gamma}_0$ in the original model. If there exists a reduction function $\boldsymbol{\theta} = R(\boldsymbol{\theta}_*, \boldsymbol{\gamma})$ and $\boldsymbol{\theta} = R(\boldsymbol{\theta}, \boldsymbol{\gamma}_0)$, the MPV can be computed via

PX-E step. Compute

$$Q\left(\boldsymbol{\theta}_*, \boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right) = \mathrm{E}_{\{\boldsymbol{\eta}_k\}|\{\widehat{\boldsymbol{\mathcal{F}}}_k\}, \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0}\left[\log p_*\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\middle|\boldsymbol{\theta}_*, \boldsymbol{\gamma}\right)\right] \tag{10}$$

PX-M step. Find

$$\left\{\boldsymbol{\theta}_*^{(t+1)}, \boldsymbol{\gamma}^{(t+1)}\right\} = \arg\max_{\boldsymbol{\theta}_*, \boldsymbol{\gamma}} Q\left(\boldsymbol{\theta}_*, \boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right) \tag{11}$$

and update $\boldsymbol{\theta}^{(t+1)} = R(\boldsymbol{\theta}_*^{(t+1)}, \boldsymbol{\gamma}^{(t+1)})$.

Since it is the ordinary EM applied to the parameter expanded complete-data model, PX-EM shares with EM its simplicity and stability. Liu et al. [34] established theoretical results to show that PX-EM can converge no slower than EM. In practice, it is observed that PX-EM converges much faster than EM. PX-EM may give better results for a fixed number of iterations because it converges faster. Different from the usual application of PX-EM to accelerate the ordinary EM, PX-EM is introduced here to handle the norm constraints on mode shapes by regarding the auxiliary parameters as vector norms.

Unlike the PX-EM algorithm that takes advantage of the particular structure of the statistical model, the P-EM algorithm utilizes the generated sequence $\{\boldsymbol{\theta}^{(t)}\}$ in the EM algorithm to make a

parabolic extrapolation to generate $\boldsymbol{\theta}^{(t+1)}$. From the perspective of fixed-point iteration, the EM algorithm, as well as the PX-EM algorithm, implicitly defines a mapping F from the parameter space onto itself such that

$$\boldsymbol{\theta}^{(t+1)} = F\left(\boldsymbol{\theta}^{(t)}\right), t \geq 0 \tag{12}$$

and we have the property that

$$\mathcal{L}\left(\boldsymbol{\theta}^{(t+1)}\right) \geq \mathcal{L}\left(\boldsymbol{\theta}^{(t)}\right), t \geq 0 \tag{13}$$

where $\mathcal{L}(\boldsymbol{\theta}) = \log p\left(\boldsymbol{\theta}\middle|\{\widehat{\boldsymbol{\mathcal{F}}}_k\}\right)$. Under usual assumptions, the last property ensures that the sequence $\{\boldsymbol{\theta}^{(t)}\}$ is convergent to some $\widehat{\boldsymbol{\theta}}$. In order to find another sequence built from $\boldsymbol{\theta}^{(0)}$ by means of F and $\mathcal{L}$ with a faster rate of convergence than $\{\boldsymbol{\theta}^{(t)}\}$, the P-EM algorithm approximates the local curvature of the surface $\left(\boldsymbol{\theta}, \mathcal{L}(\boldsymbol{\theta})\right)$ by a parabola controlled by three successively generated values of $\boldsymbol{\theta}$. The pseudo-code of the P-EM algorithm is given in Algorithm 1. Since the EM algorithm generally can move quickly its iterates to a neighborhood of a stationary point, the P-EM algorithm begins with a few iterations of EM from some starting point $\boldsymbol{\theta}^{(0)}$. Tailored to our problem, empirical procedures are proposed in Section 4.1 to determine the initial value $\boldsymbol{\theta}^{(0)}$ and the number of initial iterations n_b. In the extrapolation step, two parameters a and b need to be tuned. Good reference values of $a = 1.5$ and $b = 0.1$ are proposed in [35], and they are used in this paper.

For globally convex problems (i.e., unique optimum) that is the case for the OMA problem considered in this paper, P-EM and EM converge to the same result. The P-EM algorithm keeps the desired convergence properties of the original algorithm (stability and monotonic increase of likelihood). A remarkable gain in efficiency has been observed in a broad variety of situations where EM converges slowly [35]. In the following, we focus on developing the PX-EM algorithm for the Bayesian FFT method. The application of the P-EM algorithm is critical to improving convergence speed and it can be incorporated easily to build the final algorithm.

Algorithm 1. Pseudo-code of P-EM

1. Initialization

   Compute n_b iterates $\boldsymbol{\theta}^{(1)}$, …, $\boldsymbol{\theta}^{(n_b)}$ of EM or PX-EM from the starting value $\boldsymbol{\theta}^{(0)}$ and set $\widetilde{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}^{(n_b)}$;

2. Iterations

   $Iter = 0$; $\mathcal{L}_{old} = \mathcal{L}\left(\widetilde{\boldsymbol{\theta}}^{(0)}\right)$; $\widetilde{\boldsymbol{\theta}}^{(1)} = F(\widetilde{\boldsymbol{\theta}}^{(0)})$; $\widetilde{\boldsymbol{\theta}}^{(2)} = F(\widetilde{\boldsymbol{\theta}}^{(1)})$;

   while $Iter < Itermax$ % authorized maximum number of iterations

      $Iter = Iter + 1$; $\widetilde{\boldsymbol{\theta}}_{best} = \widetilde{\boldsymbol{\theta}}^{(2)}$; $\mathcal{L}_{best} = \mathcal{L}\left(\widetilde{\boldsymbol{\theta}}^{(2)}\right)$;

      $i = 0$; $c = 1 + a^i b$; % a and b are extrapolation parameters, chosen by the user

      $\widetilde{\boldsymbol{\theta}}_{new} = (1-c)^2\widetilde{\boldsymbol{\theta}}^{(0)} + 2c(1-c)\widetilde{\boldsymbol{\theta}}^{(1)} + c^2\widetilde{\boldsymbol{\theta}}^{(2)}$; $\mathcal{L}_{new} = \mathcal{L}\left(\widetilde{\boldsymbol{\theta}}_{new}\right)$;

      while $\mathcal{L}_{new} > \mathcal{L}_{best}$

         $\widetilde{\boldsymbol{\theta}}_{best} = \widetilde{\boldsymbol{\theta}}_{new}$; $\mathcal{L}_{best} = \mathcal{L}_{new}$;

         $i = i + 1$; $c = 1 + a^i b$;

         $\widetilde{\boldsymbol{\theta}}_{new} = (1-c)^2\widetilde{\boldsymbol{\theta}}^{(0)} + 2c(1-c)\widetilde{\boldsymbol{\theta}}^{(1)} + c^2\widetilde{\boldsymbol{\theta}}^{(2)}$;

         $\mathcal{L}_{new} = \mathcal{L}\left(\widetilde{\boldsymbol{\theta}}_{new}\right)$;

      endwhile

      if $(\mathcal{L}_{best} - \mathcal{L}_{old})/(\mathcal{L}_{best} + \mathcal{L}_{old}) < \epsilon$ % authorized convergence criterion

         STOP; % convergence achieved

      endif

      $\widetilde{\boldsymbol{\theta}}^{(0)} = \widetilde{\boldsymbol{\theta}}^{(1)}$; $\widetilde{\boldsymbol{\theta}}^{(1)} = \widetilde{\boldsymbol{\theta}}^{(2)}$; $\widetilde{\boldsymbol{\theta}}^{(2)} = F(F(\widetilde{\boldsymbol{\theta}}_{best}))$; $\mathcal{L}_{old} = \mathcal{L}_{best}$;

   endwhile

4 Expectation Maximization for Bayesian FFT method

4.1 EM for most probable value computation

Suppose we have measurement $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$. In this section we show how the EM algorithm can be applied to compute the MPV $\widehat{\boldsymbol{\theta}}$ given the probabilistic model shown in Eqns. (5) and (6). Following the procedure of EM, the key quantity in need is the Q-function, i.e., the expectation of complete-data log-likelihood, which requires the conditional joint distribution of $\widehat{\boldsymbol{\mathcal{F}}}_k$ and $\boldsymbol{\eta}_k$ given $\boldsymbol{\theta}$ and the conditional distribution of $\boldsymbol{\eta}_k$ given $\widehat{\boldsymbol{\mathcal{F}}}_k$ and $\boldsymbol{\theta}$. The derivation of these distributions is straightforward because of the property of the complex Gaussian distribution [40]. Based on the assumption of $\boldsymbol{p}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{S})$, the latent variable $\boldsymbol{\eta}_k = \boldsymbol{h}_k\boldsymbol{p}_k$ is also complex Gaussian distributed with zero-mean and a covariance matrix $\boldsymbol{H}_k = \boldsymbol{h}_k\boldsymbol{S}\boldsymbol{h}_k^*$. Accordingly, the conditional joint distribution of $\widehat{\boldsymbol{\mathcal{F}}}_k$ and $\boldsymbol{\eta}_k$ given $\boldsymbol{\theta}$, i.e., $p\left(\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k \middle| \boldsymbol{\theta}\right)$, is again a zero-mean complex Gaussian distribution with the covariance matrix

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} \boldsymbol{\Phi}\boldsymbol{H}_k\boldsymbol{\Phi}^{\mathrm{T}} + S_e\boldsymbol{I}_n & \boldsymbol{\Phi}\boldsymbol{H}_k \\ \boldsymbol{H}_k\boldsymbol{\Phi}^{\mathrm{T}} & \boldsymbol{H}_k \end{bmatrix} \tag{14}$$

Taking into account the statistical independence at different frequencies, the complete-data log-likelihood is therefore

$$L\left(\boldsymbol{\theta}\middle|\{\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\eta}_k\}\right)=\log p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\eta}_k\}\middle|\boldsymbol{\theta}\right)=-2nN_f\log\pi-nN_f\log S_e$$
$$-S_e^{-1}\sum_{k=1}^{N_f}\left[\widehat{\boldsymbol{\mathcal{F}}}_k-\boldsymbol{\Phi}\boldsymbol{\eta}_k\right]^*\left[\widehat{\boldsymbol{\mathcal{F}}}_k-\boldsymbol{\Phi}\boldsymbol{\eta}_k\right]+\sum_{k=1}^{N_f}\log|\boldsymbol{H}_k^{-1}|-\sum_{k=1}^{N_f}\boldsymbol{\eta}_k^*\boldsymbol{H}_k^{-1}\boldsymbol{\eta}_k \tag{15}$$

In this paper, all useful formulas in matrix theory are summarized in Appendix A. The derivation of Eqn. (15) can be found in Appendix B. One can show (as provided in Appendix B) that the conditional distribution $p\left(\boldsymbol{\eta}_k\middle|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}\right)$ is a complex Gaussian with the following mean and covariance:

$$\mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}}[\boldsymbol{\eta}_k]=\boldsymbol{P}_k^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\widehat{\boldsymbol{\mathcal{F}}}_k \tag{16}$$

$$\mathrm{Cov}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}}[\boldsymbol{\eta}_k]=S_e\boldsymbol{P}_k^{-1} \tag{17}$$

where we have defined $\boldsymbol{P}_k=S_e\boldsymbol{H}_k^{-1}+\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$.

Based on the above results, one can evaluate the Q-function by taking the expectation of Eqn. (15). However, if we proceed to the M step with the current Q-function, it is difficult to obtain closed-form solution for updating the mode shape $\boldsymbol{\Phi}$ because they are further subject to norm constraints. In order to resolve this issue, we apply the PX-EM algorithm by introducing the auxiliary parameter $\boldsymbol{\gamma}=\mathrm{diag}(\gamma_1,\gamma_2,\dots,\gamma_m)$ ($\gamma_i\in\mathbb{R}^+$, real positive) to obtain the augmented model

$$\widehat{\boldsymbol{\mathcal{F}}}_k=\boldsymbol{\Phi}_*\boldsymbol{\gamma}\boldsymbol{h}_k\boldsymbol{p}_k+\boldsymbol{\varepsilon}_k \tag{18}$$

where $\boldsymbol{p}_k\sim\mathcal{CN}(\boldsymbol{0},\boldsymbol{S}_*)$, $\boldsymbol{\varepsilon}_k\sim\mathcal{CN}(\boldsymbol{0},S_{e*}\boldsymbol{I}_n)$, and $\boldsymbol{\phi}_{*i}^{\mathrm{T}}\boldsymbol{\phi}_{*i}=1$ for $i=1,2,\dots,m$. By comparing the original and the augmented models, it is easy to find the following many-to-one mapping

$$\boldsymbol{\theta}=\{\boldsymbol{f},\boldsymbol{\zeta},\boldsymbol{\Phi},\boldsymbol{S},S_e\}=R(\boldsymbol{\theta}_*,\boldsymbol{\gamma})=\{\boldsymbol{f}_*,\boldsymbol{\zeta}_*,\boldsymbol{\Phi}_*,\boldsymbol{\gamma}\boldsymbol{S}_*\boldsymbol{\gamma}^{\mathrm{T}},S_{e*}\} \tag{19}$$

Moreover, the original model corresponds to the augmented model when the value of $\boldsymbol{\gamma}$ is an $(m\times m)$ identity matrix, i.e., $\boldsymbol{\gamma}_0=\boldsymbol{I}_m$. The above observations together justify why we can apply the PX-EM algorithm to infer the unknown parameters $\boldsymbol{\theta}$.

For the augmented model, we can derive the Q-function following similar steps for the original model to obtain

$$Q\left(\boldsymbol{\theta}_*,\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0\right)=-2nN_f\log\pi-nN_f\log S_{e*}-S_{e*}^{-1}Q_1\left(\boldsymbol{\Phi}_*,\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0\right)+Q_2\left(\boldsymbol{f}_*,\boldsymbol{\zeta}_*,\boldsymbol{S}_*|\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0\right) \tag{20}$$

where

$$Q_1\left(\boldsymbol{\Phi}_*,\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0\right)=\sum_{k=1}^{N_f}\widehat{\boldsymbol{\mathcal{F}}}_k^*\widehat{\boldsymbol{\mathcal{F}}}_k-2\mathrm{tr}\left[\boldsymbol{\Phi}_*\boldsymbol{\gamma}\sum_{k=1}^{N_f}\mathrm{Re}\left(\boldsymbol{w}_{1k}\widehat{\boldsymbol{\mathcal{F}}}_k^*\right)\right]+\mathrm{tr}\left[\boldsymbol{\Phi}_*\boldsymbol{\gamma}\mathrm{Re}\left(\sum_{k=1}^{N_f}\boldsymbol{w}_{2k}\right)\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\Phi}_*^{\mathrm{T}}\right] \tag{21}$$

$$Q_2\left(\boldsymbol{f}_*,\boldsymbol{\zeta}_*,\boldsymbol{S}_*|\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0\right)=\sum_{k=1}^{N_f}\log|\boldsymbol{H}_k^{-1}|-\mathrm{tr}\left[\sum_{k=1}^{N_f}\boldsymbol{H}_k^{-1}\boldsymbol{w}_{2k}\right] \tag{22}$$

$\boldsymbol{w}_{1k}=\mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}[\boldsymbol{\eta}_k]$ and $\boldsymbol{w}_{2k}=\mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}[\boldsymbol{\eta}_k\boldsymbol{\eta}_k^*]$, which are the first and second conditional moments of $\boldsymbol{\eta}_k$ given $\widehat{\boldsymbol{\mathcal{F}}}_k$, $\boldsymbol{\theta}^{(t)}$ and $\boldsymbol{\gamma}_0$; and they can be simply calculated from Eqns.

(16) and (17). Since they are evaluated at parameter values in the previous iteration, they are constant in the M-step. Note that $Q_1\left(\boldsymbol{\Phi}_*, \boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right)$ only depends on mode shapes $\boldsymbol{\Phi}_*$ and the auxiliary parameter $\boldsymbol{\gamma}$, while $Q_2\left(\boldsymbol{f}_*, \boldsymbol{\zeta}_*, \boldsymbol{S}_*|\boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right)$ involves only the remaining unknown parameters except S_{e*}. This implies that the unknown parameters can be optimized in different groups.

Once the Q-function is derived, we can then proceed to the PX-M step, i.e., to maximize $Q\left(\boldsymbol{\theta}_*, \boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right)$ w.r.t. $\boldsymbol{\theta}_*$ and $\boldsymbol{\gamma}$. Since $\boldsymbol{\Phi}_*$ and $\boldsymbol{\gamma}$ in Q_1 (and hence Q) always appear together, we can combine them into $\boldsymbol{\Phi}' = \boldsymbol{\Phi}_*\boldsymbol{\gamma}$ so that it is free from norm constraints. After $\boldsymbol{\Phi}'$ has been updated, $\boldsymbol{\Phi}_*$ and $\boldsymbol{\gamma}$ can be readily recovered by noting that the diagonal elements of $\boldsymbol{\gamma}$ are simply the norm of columns of $\boldsymbol{\Phi}'$ and $\boldsymbol{\Phi}_* = \boldsymbol{\Phi}'\boldsymbol{\gamma}^{-1}$. To obtain the optimal value of $\boldsymbol{\Phi}'$, taking the partial derivative of $Q\left(\boldsymbol{\theta}_*, \boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right)$ in Eqn. (20) w.r.t. $\boldsymbol{\Phi}'$ gives

$$\frac{\partial Q\left(\boldsymbol{\theta}_*, \boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right)}{\partial \boldsymbol{\Phi}'} = S_{e*}^{-1}\left[2\sum_{k=1}^{N_f}\mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k\boldsymbol{w}_{1k}^*\right) - 2\boldsymbol{\Phi}'\sum_{k=1}^{N_f}\mathrm{Re}(\boldsymbol{w}_{2k})\right] \tag{23}$$

Setting it to zero and solving for $\boldsymbol{\Phi}'$ yields

$$\boldsymbol{\Phi}' = \sum_{k=1}^{N_f}\mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k\boldsymbol{w}_{1k}^*\right)\left[\sum_{k=1}^{N_f}\mathrm{Re}(\boldsymbol{w}_{2k})\right]^{-1} \tag{24}$$

As can be seen, the introduction of auxiliary parameter γ in the PX-EM algorithm allows an analytical update of mode shape matrix, which can bring great saving in mathematical derivation and numerical computation compared to the Lagrange method used for the identification of closely-spaced modes [25], where the numerical optimization method has to be used. For the case of well-separated modes, since the analytical update of mode shape vector exists in the Lagrange method [23], no apparent advantage is given by the PX-EM algorithm, except that it provides a conceptually consistent solution for both cases.

Similarly, we can derive the analytical update for S_{e*} from Q in Eqn. (20) and $\boldsymbol{S}_*$ from Q_2 in Eqn. (22) as

$$S_{e*} = \frac{1}{nN_f}\mathrm{tr}\left[\sum_{k=1}^{N_f}\mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k\widehat{\boldsymbol{\mathcal{F}}}_k^*\right) - \boldsymbol{\Phi}'\sum_{k=1}^{N_f}\mathrm{Re}(\boldsymbol{w}_{2k})\,\boldsymbol{\Phi}'^{\mathrm{T}}\right] \tag{25}$$

$$\boldsymbol{S}_* = \frac{1}{N_f}\sum_{k=1}^{N_f}\boldsymbol{h}_k^{-1}\boldsymbol{w}_{2k}\boldsymbol{h}_k^{-*} \tag{26}$$

and the detailed derivation is given in Appendix B.

For $\boldsymbol{f}_*$ and $\boldsymbol{\zeta}_*$, an analytical update has not been found (as is typical), and one has to rely on numerical optimization. Since the dimension ($2m$) is not high, the MATLAB function '*fminsearch*' [41] can provide an efficient solution using the simplex search method [42].

As a remark, if the algorithm is developed using $\boldsymbol{p}_k$ instead of $\boldsymbol{\eta}_k$ as the latent variable, the only change to the updating procedure is that $\boldsymbol{w}_{1k}$ is replaced by $\boldsymbol{h}_k\mathrm{E}_{\boldsymbol{p}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}[\boldsymbol{p}_k]$ and $\boldsymbol{w}_{2k}$ is replaced by $\boldsymbol{h}_k\mathrm{E}_{\boldsymbol{p}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}[\boldsymbol{p}_k\boldsymbol{p}_k^*]\boldsymbol{h}_k^*$, where $\mathrm{E}_{\boldsymbol{p}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}[\boldsymbol{p}_k] = \boldsymbol{h}_k^{-1}\boldsymbol{P}_k^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\widehat{\boldsymbol{\mathcal{F}}}_k$ and $\mathrm{E}_{\boldsymbol{p}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}[\boldsymbol{p}_k\boldsymbol{p}_k^*] = S_e\boldsymbol{h}_k^{-1}\boldsymbol{P}_k^{-1}\boldsymbol{h}_k^{-*}$ should be calculated based on parameter values in the

previous iteration. The difference therefore only lies in the amount of information used in the previous iteration to update the parameters in the present iteration. In particular, when $\boldsymbol{p}_k$ is used as the latent variable the values of natural frequencies and damping ratios in the current iteration also affect the updated values of mode shapes through the term $\boldsymbol{h}_k$. The reverse is true for the modal force PSD matrix, however, whose value no longer depends on the values of frequencies or damping ratios in the current iteration. Intuitively, it is more preferred to use $\boldsymbol{\eta}_k$ as the latent variable because modal force PSD is more related with damping ratios than mode shapes. Numerical experiments reveal that performance is somewhat similar and so that case is not further pursued.

The other issue remaining is the selection of appropriate parameter values to initialize the iteration. This is crucial for the EM algorithm, because it is a local optimization algorithm in nature. In this paper, we generally follow the initialization procedure proposed in [25] for $\boldsymbol{f}_*$, $\boldsymbol{\zeta}_*$, $\boldsymbol{\Phi}_*$, and S_{e*} using high signal-to-noise (s/n) asymptotics results. In addition, when the s/n is high, $\boldsymbol{P}_k = S_e\boldsymbol{H}_k^{-1} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} \sim \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$, from which the second moment $\boldsymbol{w}_{2k}$ can be calculated directly from $\boldsymbol{\Phi}$. Eqn. (26) is then used to provide an initial guess for $\boldsymbol{S}_*$.

Though closed-form updates of modal parameters (except modal frequencies and damping ratios) are available in the PX-EM algorithm, experience during development of algorithm reveals that it can take a large number of iterations to converge in challenging situations, e.g., closely-spaced modes, which motivates us to further apply the P-EM algorithm to the derived PX-EM algorithm, resulting in the final algorithm denoted as P^{2}X-EM. As indicated in Algorithm 1, the PX-EM algorithm should run n_b steps to initiate. To adaptively select the parameter n_b, the convergence of the prediction error PSD S_e is checked in each iteration. This is motivated from the observation that S_e often converges quickly and can be used as an index for preliminary convergence.

4.2 Robust implementation

As an iterative algorithm, the EM algorithm is subject to accumulated numerical errors, which could destroy the monotonic convergence of the algorithm. For instance, the covariance matrices in Eqn. (17) calculated in the E step must be kept symmetric and positive semidefinite. Therefore, a robust and efficient implementation is desired to improve the performance of the proposed algorithm. For this purpose, an implementation of the PX-EM algorithm is introduced here based on the QR and Cholesky decompositions.

In the initialization, we require the square-root of a symmetric and positive semi-definite matrix, e.g.,

$$\boldsymbol{S} = \boldsymbol{S}^{1/2}\boldsymbol{S}^{*/2},\ \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} = \boldsymbol{\Phi}^{\mathrm{T}/2}\boldsymbol{\Phi}^{1/2} \tag{27}$$

which can be done by the Cholesky decomposition. Then, we apply the QR decomposition for the PX-E step,

$$\begin{bmatrix} \boldsymbol{\Phi}^{1/2} & \boldsymbol{\Phi}^{-\mathrm{T}/2}\boldsymbol{\Phi}^{\mathrm{T}}\widehat{\boldsymbol{\mathcal{F}}}_k \\ \sqrt{S_e}\boldsymbol{S}^{-1/2}\boldsymbol{h}_k^{-1} & \boldsymbol{0}_{m\times 1} \end{bmatrix} = \widetilde{\boldsymbol{Q}} \begin{bmatrix} \widetilde{\boldsymbol{\mathcal{R}}}_{11} & \widetilde{\boldsymbol{\mathcal{R}}}_{12} \\ \boldsymbol{0}_m & \widetilde{\boldsymbol{\mathcal{R}}}_{22} \end{bmatrix} \tag{28}$$

Taking advantage of the unitary nature of $\widetilde{\boldsymbol{Q}}$, one can verify that

$$\mathrm{E}\left[\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)}\right]=\widetilde{\boldsymbol{\mathcal{R}}}_{11}^{-1}\widetilde{\boldsymbol{\mathcal{R}}}_{12},\ \mathrm{Cov}\left[\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)}\right]=S_e\widetilde{\boldsymbol{\mathcal{R}}}_{11}^{-1}\widetilde{\boldsymbol{\mathcal{R}}}_{11}^{-\mathrm{T}}, \tag{29}$$

For the PX-M step, we perform the Cholesky decomposition

$$\begin{bmatrix}\boldsymbol{\mathcal{H}} & \boldsymbol{\mathcal{J}}^{\mathrm{T}} & \mathbf{0}_m\\ \boldsymbol{\mathcal{J}} & \boldsymbol{\mathcal{K}} & \mathbf{0}_{n\times m}\\ \mathbf{0}_m & \mathbf{0}_{m\times n} & \boldsymbol{\mathcal{S}}\end{bmatrix}=\begin{bmatrix}\tilde{\boldsymbol{L}}_{11} & \mathbf{0}_{m\times n} & \mathbf{0}_m\\ \tilde{\boldsymbol{L}}_{21} & \tilde{\boldsymbol{L}}_{22} & \mathbf{0}_{n\times m}\\ \mathbf{0}_m & \mathbf{0}_{m\times n} & \tilde{\boldsymbol{L}}_{33}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{L}}_{11}^{\mathrm{T}} & \tilde{\boldsymbol{L}}_{21}^{\mathrm{T}} & \mathbf{0}_m\\ \mathbf{0}_{n\times m} & \tilde{\boldsymbol{L}}_{22}^{\mathrm{T}} & \mathbf{0}_{n\times m}\\ \mathbf{0}_m & \mathbf{0}_{m\times n} & \tilde{\boldsymbol{L}}_{33}^{*}\end{bmatrix} \tag{30}$$

where

$$\begin{aligned}&\boldsymbol{\mathcal{H}}=N_f^{-1}\textstyle\sum_{k=1}^{N_f}\mathrm{Re}(\boldsymbol{w}_{2k}),\ \boldsymbol{\mathcal{J}}=N_f^{-1}\sum_{k=1}^{N_f}\mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k\boldsymbol{w}_{1k}^{*}\right),\\ &\boldsymbol{\mathcal{K}}=N_f^{-1}\textstyle\sum_{k=1}^{N_f}\mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k\widehat{\boldsymbol{\mathcal{F}}}_k^{*}\right),\ \boldsymbol{\mathcal{S}}=N_f^{-1}\sum_{k=1}^{N_f}\boldsymbol{h}_k^{-1}\boldsymbol{w}_{2k}\boldsymbol{h}_k^{-*}\end{aligned} \tag{31}$$

By equating submatrices, we have

$$\boldsymbol{\Phi}'=\tilde{\boldsymbol{L}}_{21}\tilde{\boldsymbol{L}}_{11}^{-1},\ S_{e*}=n^{-1}\mathrm{tr}[\tilde{\boldsymbol{L}}_{21}\tilde{\boldsymbol{L}}_{21}^{\mathrm{T}}],\ \boldsymbol{S}_{*}=\tilde{\boldsymbol{L}}_{33}\tilde{\boldsymbol{L}}_{33}^{*} \tag{32}$$

The above QR and Cholesky decompositions provide an efficient and robust way to perform the PX-EM algorithm. The main procedures are summarized in Algorithm 2.

Algorithm 2. The robust PX-EM algorithm for MPV computation in Bayesian FFT method

1. Initialization

   Set $\boldsymbol{f}^{(0)}$, $\boldsymbol{\zeta}^{(0)}$, $S_e^{(0)}$, $\boldsymbol{\Phi}^{(0)}$, $\boldsymbol{S}^{(0)}$ and $\boldsymbol{\gamma}_0=\boldsymbol{I}_m$

   $\boldsymbol{\Phi}^{1/2,(0)}=\mathrm{chol}(\boldsymbol{\Phi}^{(0),\mathrm{T}}\boldsymbol{\Phi}^{(0)})$; $\boldsymbol{S}^{-1/2,(0)}=\mathrm{chol}(\boldsymbol{S}^{(0),-1})$,

2. PX-EM iteration

   $t=0$;

   while $t<Itermax1$ (authorized maximum number of iterations)

   2.1 Robust PX-E step

   for $k=1$ to N_f

   $$\widetilde{\boldsymbol{\mathcal{R}}}=\mathrm{qr}\left(\begin{bmatrix}\boldsymbol{\Phi}^{1/2,(t)} & \boldsymbol{\Phi}^{-1/2,(t)}\boldsymbol{\Phi}^{(t),\mathrm{T}}\widehat{\boldsymbol{\mathcal{F}}}_k\\ \sqrt{S_e^{(t)}}\boldsymbol{S}^{-1/2,(t)}\boldsymbol{h}_k^{(t),-1} & \mathbf{0}_{m\times 1}\end{bmatrix}\right)$$

   $$\widetilde{\boldsymbol{\mathcal{R}}}'=\widetilde{\boldsymbol{\mathcal{R}}}_{11}^{-1}\begin{bmatrix}\sqrt{S_e^{(t)}}\boldsymbol{I}_m & \widetilde{\boldsymbol{\mathcal{R}}}_{12}\end{bmatrix};$$

   $\boldsymbol{w}_{1k}=\widetilde{\boldsymbol{\mathcal{R}}}'(:,end)$; $\boldsymbol{w}_{2k}=\widetilde{\boldsymbol{\mathcal{R}}}'\widetilde{\boldsymbol{\mathcal{R}}}'^{\mathrm{T}}$;

   endfor

   2.2 Robust PX-M step

   $$\tilde{\boldsymbol{L}}=\mathrm{chol}\left(\begin{bmatrix}\boldsymbol{\mathcal{H}} & \boldsymbol{\mathcal{J}}^{\mathrm{T}} & \mathbf{0}_m\\ \boldsymbol{\mathcal{J}} & \boldsymbol{\mathcal{K}} & \mathbf{0}_{n\times m}\\ \mathbf{0}_m & \mathbf{0}_{m\times n} & \boldsymbol{\mathcal{S}}\end{bmatrix}\right);$$

   $\boldsymbol{\Phi}'^{(t+1)}=\tilde{\boldsymbol{L}}_{21}\tilde{\boldsymbol{L}}_{11}^{-1}$; $\boldsymbol{\gamma}_{*}^{(t+1)}=\mathrm{diag}\left(\left|\boldsymbol{\Phi}'^{(t+1)}\right|_{\mathrm{col}}\right)$; $\boldsymbol{\Phi}_{*}^{(t+1)}=\boldsymbol{\Phi}'^{(t+1)}\boldsymbol{\gamma}_{*}^{(t+1),-1}$;

$S_{e*}^{(t+1)} = n^{-1}\mathrm{tr}[\tilde{\boldsymbol{L}}_{21}\tilde{\boldsymbol{L}}_{21}^{\mathrm{T}}]$; $\boldsymbol{S}_{*}^{-1/2,(t+1)} = \tilde{\boldsymbol{L}}_{33}^{-1}$;
Obtain $\boldsymbol{f}_{*}^{(t+1)}$ and $\boldsymbol{\zeta}_{*}^{(t+1)}$ using simplex search method;
$\boldsymbol{f}^{(t+1)} = \boldsymbol{f}_{*}^{(t+1)}$; $\boldsymbol{\zeta}^{(t+1)} = \boldsymbol{\zeta}_{*}^{(t+1)}$; $S_e^{(t+1)} = S_{e*}^{(t+1)}$;
$\boldsymbol{\Phi}^{(t+1)} = \boldsymbol{\Phi}_{*}^{(t+1)}$; $\boldsymbol{\Phi}^{1/2,(t+1)} = \mathrm{chol}\left(\boldsymbol{\Phi}^{(t+1),\mathrm{T}}\boldsymbol{\Phi}^{(t+1)}\right)$;
$\boldsymbol{S}^{-1/2,(t+1)} = \boldsymbol{\gamma}_{*}^{(t+1),-1}\boldsymbol{S}_{*}^{-1/2,(t+1)}$; $\boldsymbol{\gamma}_0 = \boldsymbol{I}_m$
$t = t + 1$;
if $\left|S_{e*}^{(t)} - S_{e*}^{(t-1)}\right|/(S_{e*}^{(t)} + S_{e*}^{(t-1)}) < \epsilon_1$ (authorized convergence criterion)
STOP; (convergence achieved)
endif
endwhile

5 Empirical Studies

In order to validate the performance of the proposed EM algorithm for the MPV estimation in the Bayesian FFT method, empirical studies are presented here with synthetic, laboratory, and field test data. The synthetic data example is used to evaluate the consistency of algorithm. The second (laboratory) and the third (field) examples illustrate performance in controlled and operation condition, respectively. In addition to the EM algorithm, the results calculated by the direct method [23,25] are also provided for comparison. Since we do not develop the EM algorithm to compute posterior covariance matrix, all the uncertainty information is calculated based on the direct differentiation algorithm proposed in Ref. [26].

5.1 Synthetic data

Consider synthetic data generated with three closely-spaced modes and three measurement channels. The data are generated at 100 Hz by

$$\hat{\boldsymbol{y}}_j = \boldsymbol{\phi}_1\ddot{\eta}_1(t_j) + \boldsymbol{\phi}_2\ddot{\eta}_2(t_j) + \boldsymbol{\phi}_3\ddot{\eta}_3(t_j) + \boldsymbol{e}(t_j) \tag{33}$$

where $\ddot{\eta}_i(t)$ $(i = 1,2,3)$ is the modal acceleration response, satisfying the governing equation

$$\ddot{\eta}_i(t) + 2\zeta_i\omega_i\dot{\eta}_i(t) + \omega_i^2\eta_i(t) = p_i(t) \tag{34}$$

with natural frequencies $\omega_i = 2\pi f_i\, rad/sec$, $f_1 = 0.98\, Hz$, $f_2 = 1.00\, Hz$, and $f_3 = 1.02\, Hz$; and damping ratios $\zeta_1 = 0.8\%$, $\zeta_2 = 1.0\%$, and $\zeta_3 = 1.2\%$; The modal excitations $p_i(t)$ $(i = 1,2,3)$ are modeled as stationary Gaussian white noises with PSD $S_{11} = S_{22} = S_{33} = 1\ (\mu g)^2/Hz$ $(1\ \mu g = 9.81 \times 10^{-6}\ m/sec^2)$, cross PSD $S_{12} = 0.5e^{\mathbf{i}\pi/4}\ (\mu g)^2/Hz$ and zeros elsewhere;

$$\boldsymbol{\phi}_1 = [1,2,2]^{\mathrm{T}}/3,\ \boldsymbol{\phi}_2 = [2,1,-2]^{\mathrm{T}}/3,\ \boldsymbol{\phi}_3 = [1,-2,2]^{\mathrm{T}}/3 \tag{35}$$

$\boldsymbol{e}(t)$ represents the measurement error, modeled by a stationary Gaussian white noise process with PSD $S_e\boldsymbol{I}_3$ and $S_e = 10\ (\mu g)^2/Hz$. The resulting modal signal-to-noise (s/n) ratio, defined as the PSD ratio of response to noise at the natural frequency, is around $\gamma_i = S_{ii}/4S_e\zeta_i^2 = 250$ for all modes. This quality of data can be readily achieved in typical ambient vibration tests. The PSD and singular value (SV) spectrum calculated using $5000\ sec$ of data is shown in Figure 1. The SV

spectrum plots the eigenvalues of the PSD matrix, giving a rough idea of natural frequencies and the quality of data.

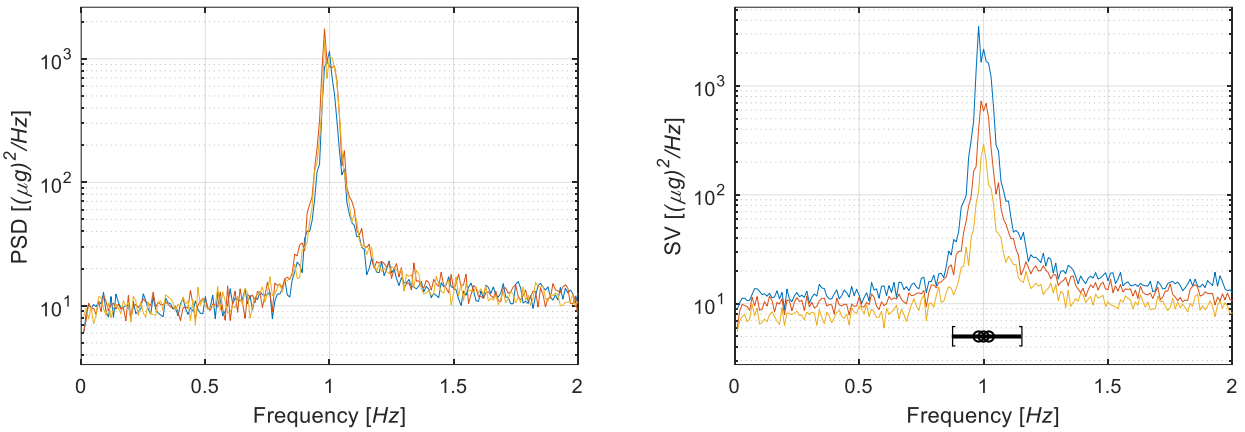


Figure 1: PSD and SV spectrum, synthetic data.
Bracket: frequency band used for modal identification.

By observing the SV spectrum of Figure 1, the frequency band is selected to be $[0.87,1.15]\ Hz$ where the top SV spectrum line (which reflects modal response) almost meets the second top line in the flat region (which reflects prediction error). See also Examples 12.2 and 12.3 on p.378-381 of [3] for investigation on the choice of bandwidth on results. Both the EM and direct methods are applied to identify the modal parameters. The results are summarized in Table 1 for the data duration of $5000\ sec$. As can be seen, the identified parameters are close to their exact values. The modal frequencies have a coefficient of variation (c.o.v.) of only about 0.07%. The damping ratios have a c.o.v. of about 7%, which is similar to the c.o.v. of the modal force PSDs. Since the MPVs of S_{23} and S_{31} are small, large c.o.v. values are yielded. The most probable and the exact mode shapes are close, evidenced by the modal assurance criterion (MAC) ≈ 1. The mode shape c.o.v., defined as the trace of posterior covariance matrix [43], is of the same order of magnitude. The EM and direct method give identical results, counter-checking each other. Since we applied a strict criterion to determine the convergence of the algorithms, i.e., the maximal relative change of modal parameters in two successive iterations is less than 10^{-5}, the two methods yield the same values up to the third decimal place.

Table 1: Modal identification results, synthetic data

Parameter		Mode 1	Mode 2	Mode 3
Frequency f_i [Hz]	Exact	0.980	1.000	1.020
	EM	0.980 (0.056)	1.000 (0.066)	1.020 (0.069)
	Direct	0.980 (0.056)	1.000 (0.066)	1.020 (0.069)
Damping ratio ζ_i [%]	Exact	0.800	1.000	1.200
	EM	0.828 (7.1)	1.038 (6.4)	1.156 (6.6)
	Direct	0.828 (7.1)	1.038 (6.4)	1.156 (6.6)
Mode shape MAC	EM	1.000 (3.3)	0.999 (4.7)	1.000 (3.6)
	Direct	1.000 (3.3)	0.999 (4.7)	1.000 (3.6)
Modal force PSD S_{ii} $[(\mu g)^2/Hz]$	Exact	1.000	1.000	1.000
	EM	0.985 (5.2)	1.015 (6.1)	1.006 (5.5)
	Direct	0.985 (5.2)	1.015 (6.1)	1.006 (5.5)
	Exact	0.350	0.000	0.000

Cross PSD, real part of S_{12}, S_{23}, S_{31} $[(\mu g)^2/Hz]$	EM	0.349 (2.7)	0.084 (46)	0.011 (839)
	Direct	0.349 (2.7)	0.084 (46)	0.011 (839)
Cross PSD, imaginary part of S_{12} , S_{23} , S_{31} $[(\mu g)^2/Hz]$	Exact	0.350	0.000	0.000
	EM	0.347 (0.67)	0.044 (44)	-0.057 (27)
	Direct	0.347 (0.67)	0.044 (44)	-0.057 (27)
Prediction error PSD S_e $[(\mu g)^2/Hz]$	Exact	10.00		
	EM	10.36 (6.2)		
	Direct	10.36 (6.2)		

Note: MPV with c.o.v. (coefficient of variation) in parenthesis (units: %) are used; the MAC (modal assurance criterion) is calculated based on the MPV of identified mode shapes and the exact ones.

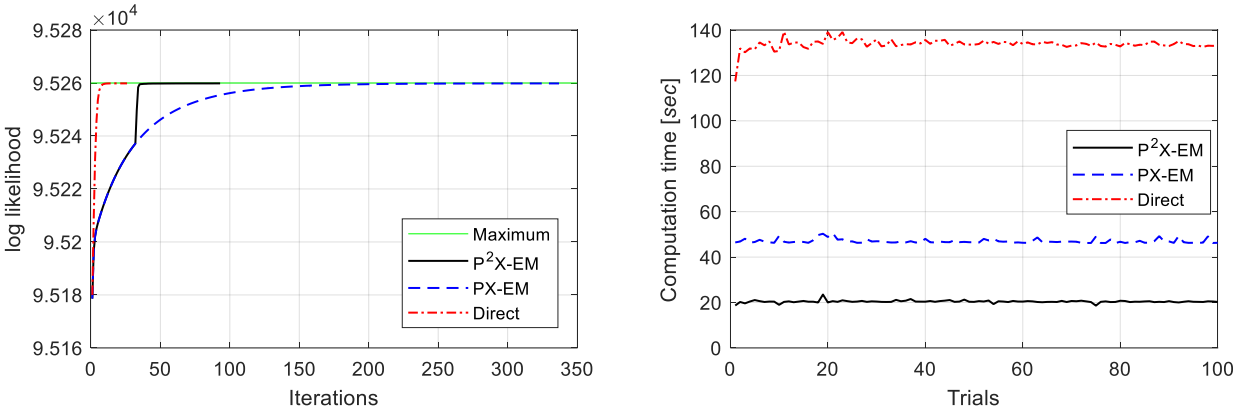


Figure 2: Performance comparison of the EM and direct methods, synthetic data

The iteration process and computation time of two variants of EM algorithms and the direct method are provided in Figure 2 for a further comparison. For this example, the P^{2}X-EM algorithm takes 93 iterations to converge, which is more than the direct method (25 iterations). However, it is computationally faster than the direct method, because closed-form expressions are available in updating unknown parameters except frequencies and damping ratios. P^{2}X-EM and the direct method spend respectively 20.4 sec and 133.8 sec averaged over 100 successive trials on a Digital Storm laptop with Intel® Core™ i7 CPU @2.50 GHz and RAM 16.0 GB. The PX-EM algorithm takes 337 iterations and 47.0 sec to converge. A large saving has been achieved both in the number of iterations and computation, demonstrating the benefit of the P^{2}X-EM algorithm.

To illustrate the consistency of the algorithm, as an example, the identified results of the first mode are shown in Figure 3 for different data lengths. The dashed line denotes the exact value of the modal parameter that generated the data. The identified result for each data length is shown with a dot at the posterior MPV and an error bar covering two posterior standard deviations. It is seen that as the data length increases the error bar generally narrows. There is no evidence of bias, as reflected by the fact that the error bar covers the exact value regardless of data length.

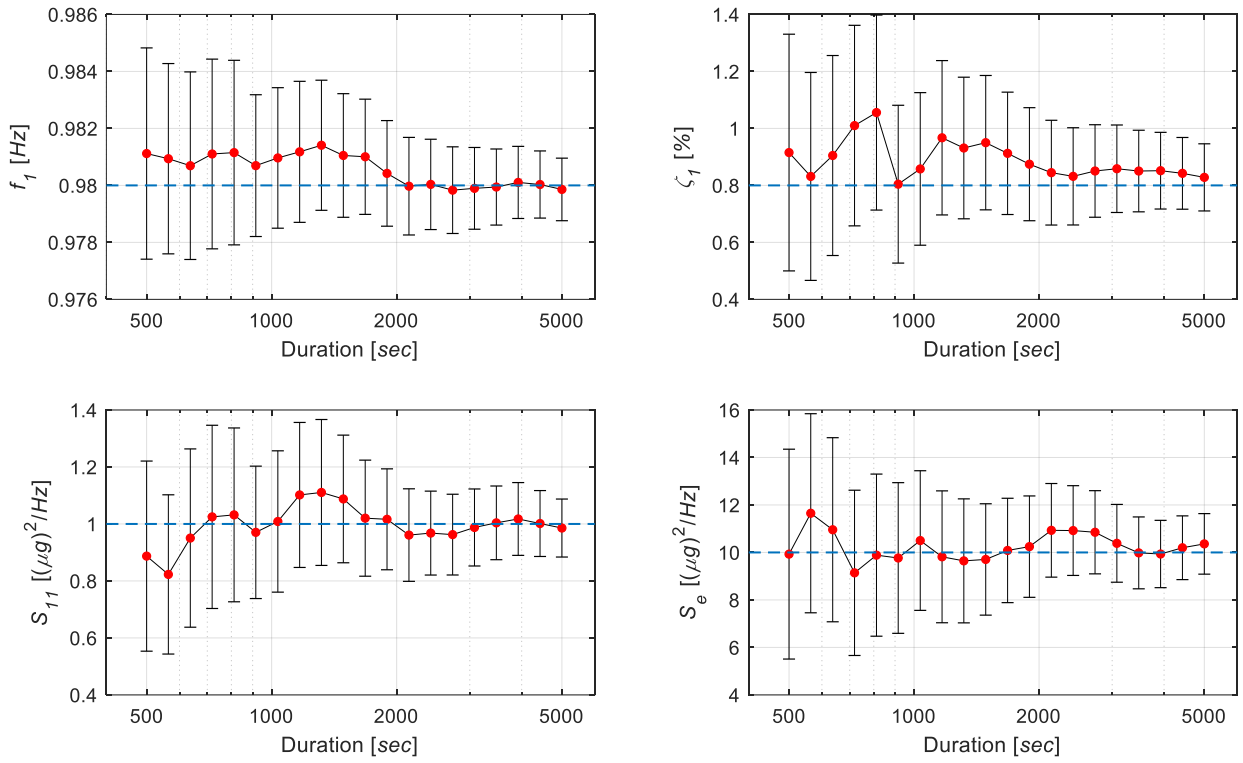
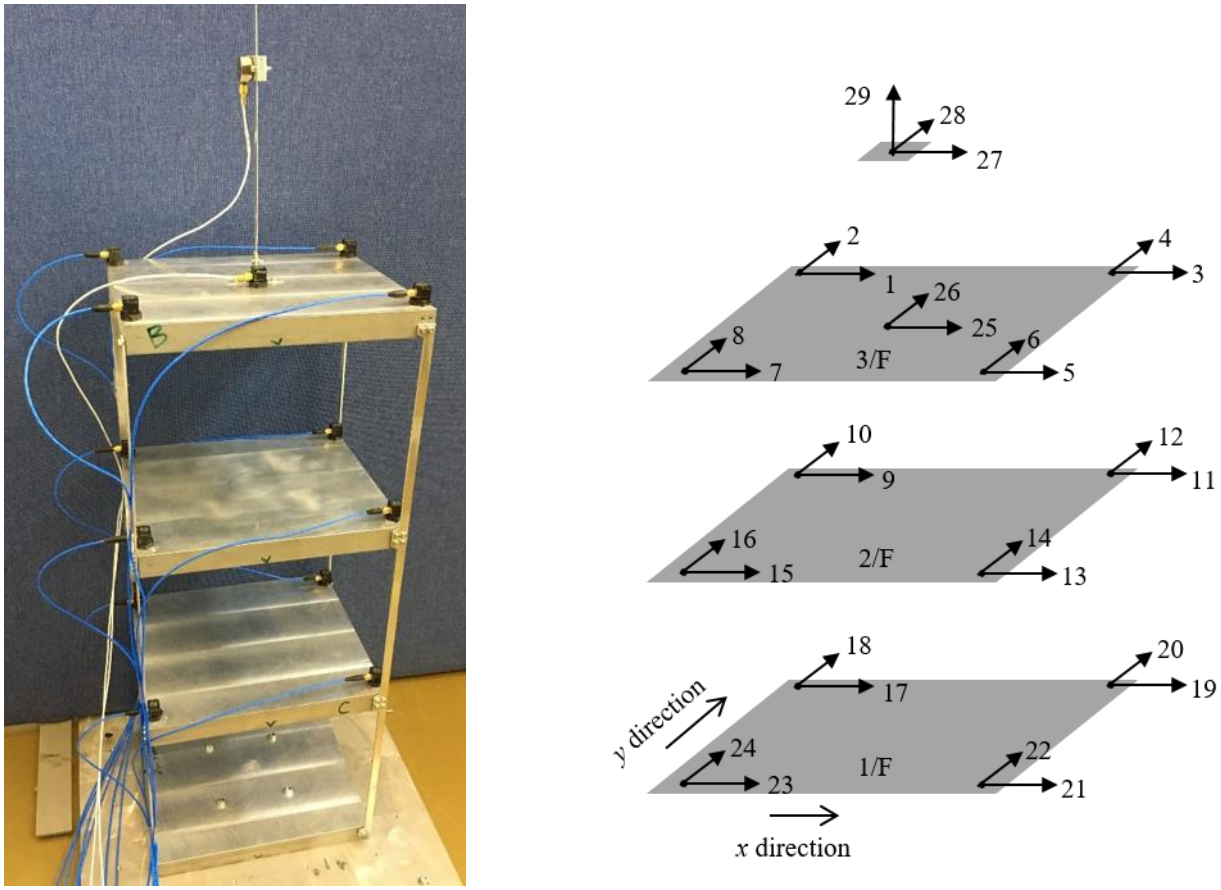


Figure 3: Identified first modal parameters, synthetic data.
Dot: MPV; error bar: ±two standard deviations; dashed line: actual value.

5.2 Laboratory data

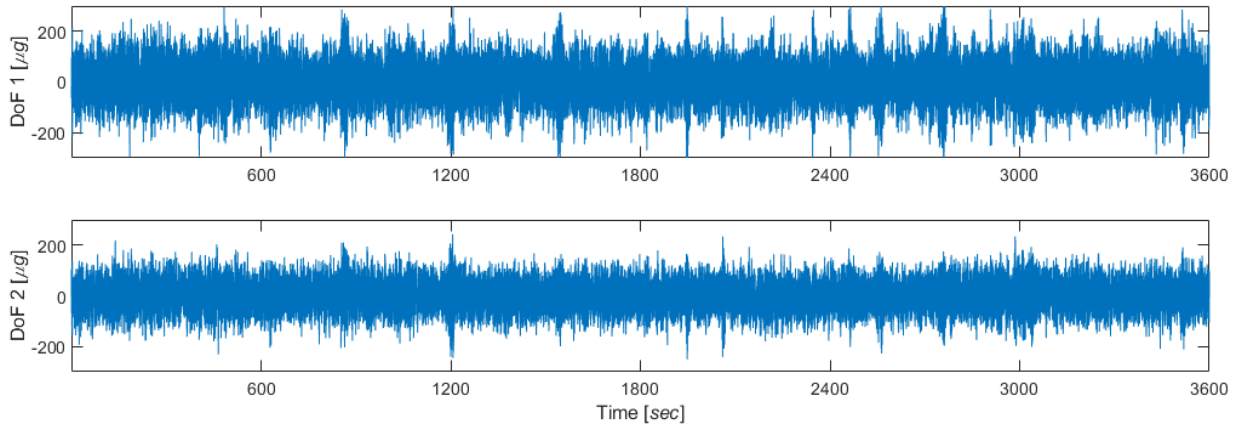
An example with laboratory data is next presented to illustrate the performance of the proposed algorithm on identifying closely-spaced modes with experimental data. Ambient vibration data was collected from a three-story aluminium frame with an added mass that models the presence of the tuned mass damper (TMD), shown in Figure 4a. The main frame measures 30 cm × 30 cm in plan with a uniform story height of 25 cm. The height of the TMD is 27 cm. Due to the similar layout of mass and stiffness in the two horizontal directions and the existence of TMD, multiple closely-space modes are expected. The four corners of each floor and the bottom of TMD were instrumented with a biaxial piezoelectric accelerometer measuring the horizontal *x*- and *y*-directions. The TMD was instrumented with a triaxial piezoelectric accelerometer measuring the vibration in the *x*-, *y*- and *z*- directions, giving a total of 29 measured DoFs (Figure 4b). Ambient data was recorded for 3600 sec at 2048 Hz and then decimated to 256 Hz for analysis. The measured time history at DoFs 1 and 2 are shown in Figure 5a. The PSD and SV spectra of data are shown in Figure 5b, which also depicts the selected bands, initial guess of frequencies and number of modes in case where more than 1 mode is recognized. By reading peaks in the spectra, six frequency bands are selected for frequencies less than 35 Hz, including 2 bands with 3 modes, 1 band with 2 modes and 3 bands with a single mode, yielding 11 modes in total.



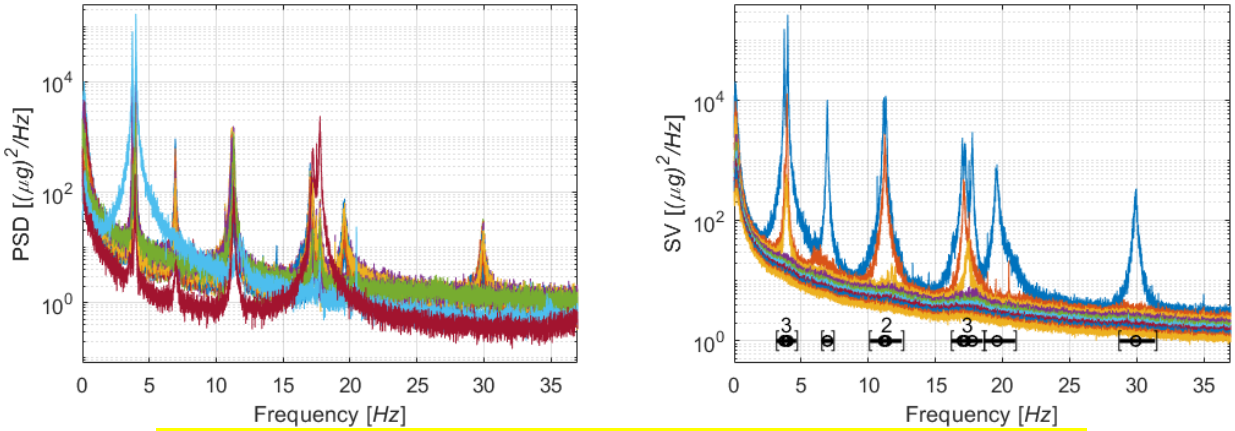
a) Photo b) Measured DoFs

Figure 4: Laboratory shear frame.

In the modal analysis, we first separated the full-length data into 6 non-overlapping segments of equal length, i.e., each with a duration of $600\ sec$. This is more than 2000 cycles of fundamental mode, typically long enough for a reliable estimation. The identified 11 modes by the EM algorithm are listed in Table 2 for the first data set. The MPVs of identified mode shapes are plotted in Figure 6 and named according to their nature, e.g., TX1 for the first translational mode along the *x*-direction; R1 for the first torsional mode. Based on lumped mass structural dynamics, the main frame theoretically has nine modes, comprising three translational modes in the *x*- and *y*-direction, and three rotational modes. The bending of the TMD in the *x*- and *y*-direction introduces two new modes, whose frequencies are close to the first and the third pairs of translational modes, respectively. The frequency is associated with small posterior uncertainty, indicating a reliable estimation. The posterior c.o.v. tends to decrease with the mode number. This is partly because the effective data length (data duration/natural period) is longer for higher modes with higher frequencies. All damping ratios except for the 10th mode are less than 0.5%, and their posterior c.o.v.s are all around or less than 10%. The MPVs of the modal force of the first three modes are much larger than the remaining ones. The MPV of prediction error PSD tends to decrease with the mode number. This is consistent with the spectra in Figure 5b. By observing the mode shapes, the translation of TMD is much larger than that of the main frame for modes TX1/TX2 and TY3/TY4, which is typical. We have not shown the results identified by the direct method, because it gives almost identical values as the EM algorithm.



a) Measured time history at DoFs 1 and 2 (detrended)



b) PSD and SV spectrum (Bracket: frequency bands used in the Bayesian FFT method)

Figure 5: Measured data, laboratory shear frame.

Table 2: Modal identification results, laboratory shear frame.

Mode	Frequency		Damping ratio		Modal force PSD		Prediction error PSD		Mode shape	
	MPV [Hz]	c.o.v. [‰]	MPV [%]	c.o.v. [%]	MPV [$(\mu g)^2/Hz$]	c.o.v. [%]	MPV [$(\mu g)^2/Hz$]	c.o.v. [%]	Characteristics	c.o.v. [%]
1	3.734	0.54	0.418	5.5	14.25	4.1	7.88	0.65	TX1	1.1
2	3.926	0.56	0.427	8.9	3.47	6.8	7.88	0.65	TY1	6.9
3	3.995	0.55	0.425	5.7	16.68	6.3	7.88	0.65	TX2	1.8
4	6.947	0.45	0.462	10	0.77	5.3	4.71	0.85	R1	1.1
5	11.141	0.34	0.448	7.8	0.81	3.0	2.59	0.51	TX3	2.1
6	11.308	0.28	0.310	9.3	0.53	3.0	2.59	0.51	TY2	6.4
7	16.995	0.27	0.421	6.7	0.13	3.6	1.85	0.52	TX4	2.1
8	17.222	0.25	0.358	7.3	0.10	3.5	1.85	0.52	TY3	2.7
9	17.739	0.22	0.261	8.4	0.08	3.3	1.85	0.52	TY4	1.5
10	19.612	0.38	0.850	5.2	0.17	3.6	1.68	0.51	R2	1.2
11	29.920	0.21	0.420	5.5	0.02	3.4	1.05	0.47	R3	1.5

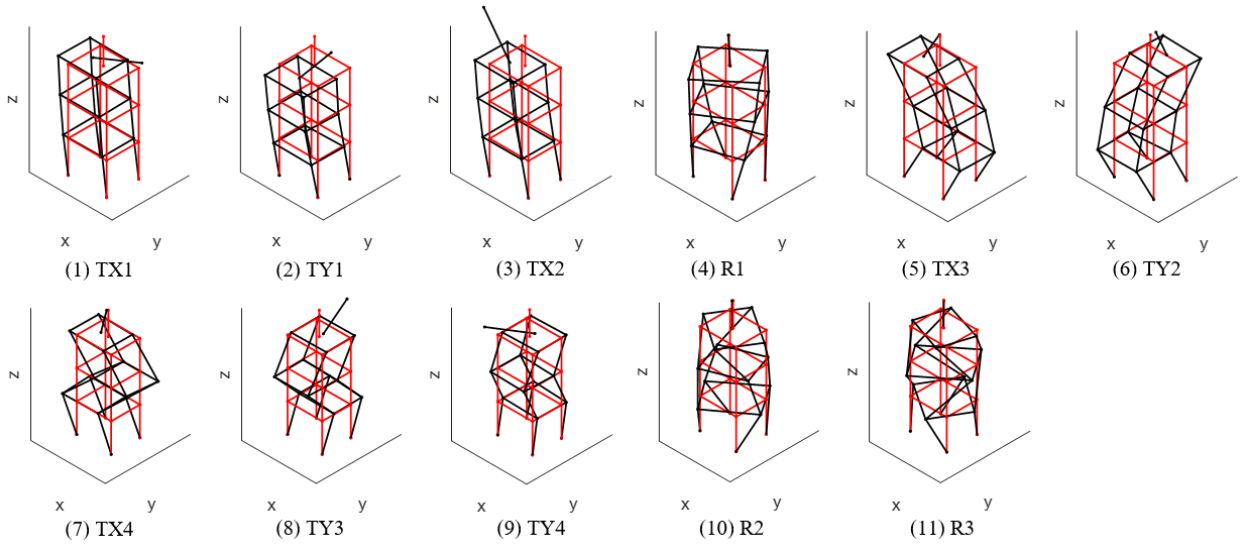


Figure 6: Identified modal shapes, laboratory shear frame.

The posterior c.o.v. tells the identification uncertainty of parameters given each data set but it does not tell the potential variability of results over different data sets, which can be due to change in the structure, environment, or merely modeling error. In order to investigate the variability of identified modal parameters for different data sets, we provide the results for modes 2 and 3 as an example, shown in Figure 7. The frequency, damping ratio and prediction error PSD show small variability. Their MPVs do not change much and the posterior uncertainties are basically the same. The modal force PSD of the 3rd mode fluctuates over different sets, indicating the change of excitation with time.

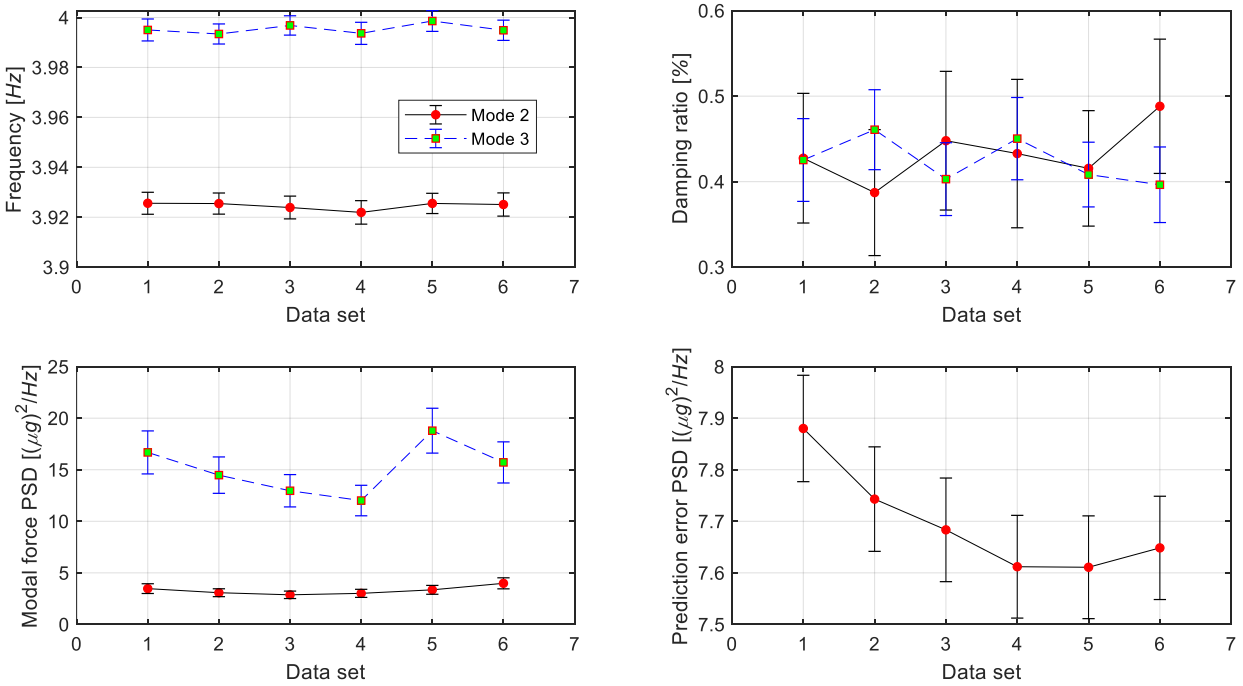
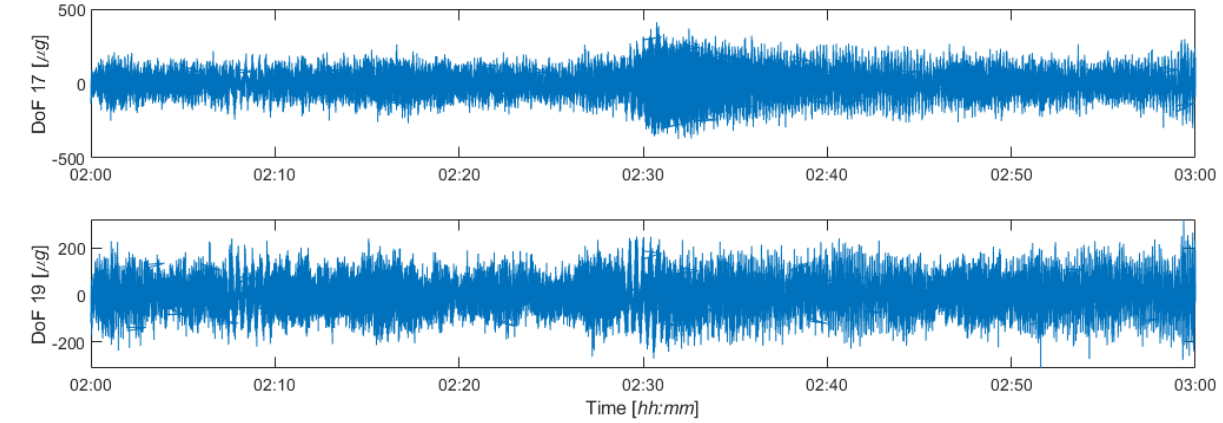


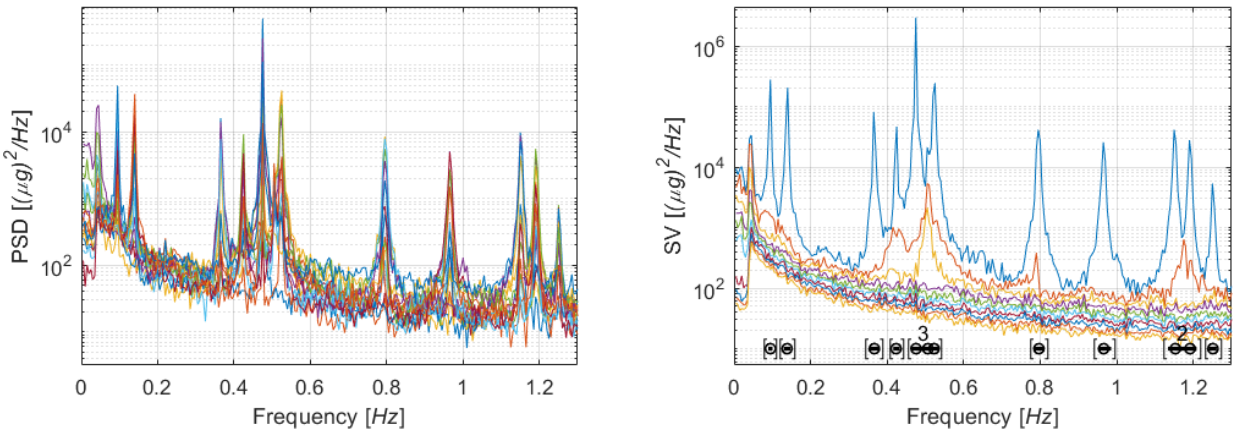
Figure 7: Identified modal parameters for different data sets, laboratory shear frame.

5.3 Field data

The third example is the Guangzhou New TV Tower (GNTVT), which is a structural health monitoring benchmark building for high-rise structures [44,45]. In the benchmark study, 24-hour field measurements of the structural acceleration time histories and the corresponding ambient conditions (temperature and wind) were provided. Twenty uni-axial accelerometers were installed along the tower, to measure the structural dynamic response. Figure 8 shows two typical time series measured from 2:00 to 3:00 on Jan. 20, 2010, as well as the corresponding PSD and SV spectrum. The selected frequency bands are also marked in the SV spectra, including one three-mode band, one two-mode band and seven single-mode band. Various methods have been applied to identify the modal parameters of GNTVT in the literature, e.g., stochastic subspace identification method [46,47], enhanced frequency domain decomposition [46,47], Bayesian spectral density approach [48] and Bayesian FFT method [49]. For the objectives in this paper, we do not intend to compare our method with these approaches, so their results are not provided in this paper. Generally speaking, our results are consistent with these previous studies. Interested readers may refer to the cited papers. As for the difference with the study in [49], multi-mode frequency bands are used in the current analysis, while only single-mode bands are applied in [49].



a) Measured time history at two locations



b) PSD and SV spectrum (Bracket: frequency bands used in the Bayesian FFT method)

Figure 8: Measured data, Guangzhou New TV Tower (Jan-20-2010).

Before the analysis, we first divide each one-hour data into non-overlapping time windows at 20-min intervals, which is about 1300 cycles of fundamental mode, and found to be long enough

for a good estimation. Though longer durations could reduce the posterior uncertainty of identified parameters, it also increases the risk of violating the stationary assumption. Regardless of whether this is explicitly recognized or not, this risk applies to many OMA methods as the assumption is quite common.

Table 3 shows the modal parameters of the first twelve modes, identified by the EM algorithm, for the first 20-min of the data given in Figure 8a, i.e., 2:00-2:20. The fundamental mode is less than 0.1Hz, representing a highly flexible structure, because the GNTVT is as high as $610m$. Again, the frequency is identified with a good precision, reflected by the small posterior c.o.v.. Damping ratios are all less than 1%, which may be due to the small amplitude in the ambient vibration condition. The posterior uncertainties associated with damping ratios are much larger than those of frequencies, implying that damping ratios are more difficult to identify precisely. The modal force PSD represents the modal participation of external forces. MPVs of of the first two modes are significantly greater than the remaining ones. The prediction error includes the measurement error and possibly modeling error due to the truncation of neighbored modes in band selection. The decreasing trend of the MPV of prediction error PSD is consistent with the SV spectrum shown in Figure 8b. In addition, the mode shapes with their characteristics are provided in Figure 9 in terms of their projections in the *x*- and *y*-direction. Ten translatonal modes and two rotational modes are recognized, and the rotational mode shapes have relatively larger posterior c.o.v.s. Similar to the last example, the results identified by the direct method are not given, because they are identical to the values in the EM algorithm.

Table 3: Modal identification results, Guangzhou New TV Tower.

Mode	Frequency		Damping ratio		Modal force PSD		Prediction error PSD		Mode shape	
	MPV $[Hz]$	c.o.v. [‰]	MPV [%]	c.o.v. [%]	MPV $[(\mu g)^2/Hz]$	c.o.v. [%]	MPV $[(\mu g)^2/Hz]$	c.o.v. [%]	Characteristics	c.o.v. [%]
1	0.094	2.5	0.441	75	131.01	27	276.57	6.1	TY1	3.0
2	0.139	2.6	0.618	45	106.36	24	173.22	5.3	TX1	2.5
3	0.366	1.4	0.439	34	6.54	20	52.37	4.2	TY2	2.9
4	0.423	1.4	0.480	36	2.69	29	65.09	5.3	TX2	5.0
5	0.475	0.65	0.141	40	20.48	11	34.57	3.0	TY3	0.77
6	0.506	0.82	0.246	35	1.36	14	34.57	3.0	R1	7.9
7	0.523	0.87	0.272	25	5.40	13	34.57	3.0	TX3	2.9
8	0.796	0.68	0.244	30	2.54	16	32.74	3.5	TY4	1.8
9	0.965	0.49	0.156	33	0.91	15	22.86	3.5	TX4	1.8
10	1.152	0.34	0.096	36	0.69	12	17.36	2.8	TY5	1.5
11	1.192	0.41	0.131	31	0.44	13	17.36	2.8	TX5	2.5
12	1.251	0.51	0.181	41	0.13	38	24.18	4.9	R2	4.9

To investigate the temporal variability of modal parameters, the EM algorithm was applied to all 72 data sets. The identification results are plotted in Figure 10, with solid line representing the MPV and shaded area covering two standard deviations. The frequencies change slightly with time, while the damping ratios have a larger variation but still are in the same order of magnitude. The negative values of damping ratios in the plot is immaterial, merely due to the Gausssian distribution approximation and the large c.o.v.. By comparing the sample c.o.v. and the posterior c.o.v. (median of c.o.v. of all 72 data sets) in Table 4, the posterior uncertainty is consistent with the ensemble variability of their MPVs and therefore the Bayesian and frequentist perspectives roughly agree [50]. In addition, the posterior c.o.v. tends to be smaller than the sample c.o.v., because the sample c.o.v. has incorporated the uncertainties due to the environmal effects, e.g., temperature, wind.

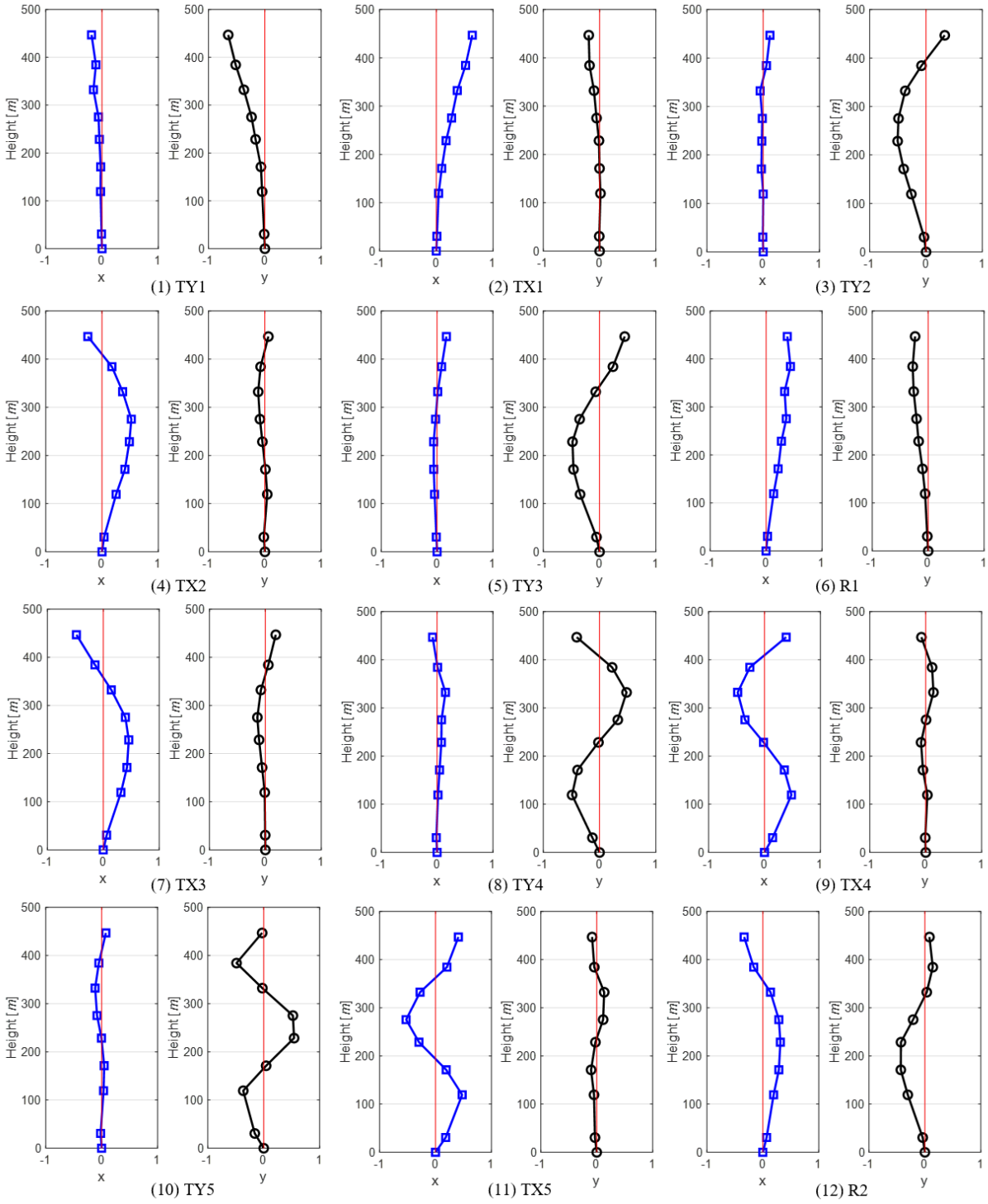
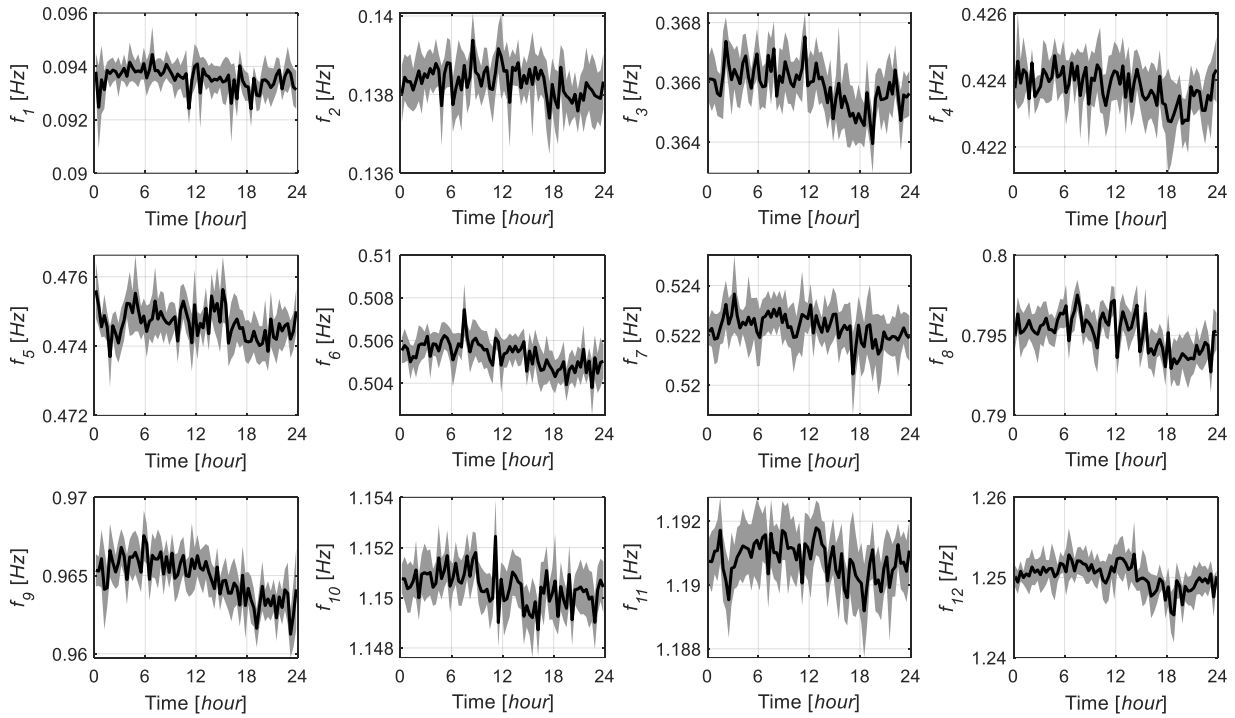
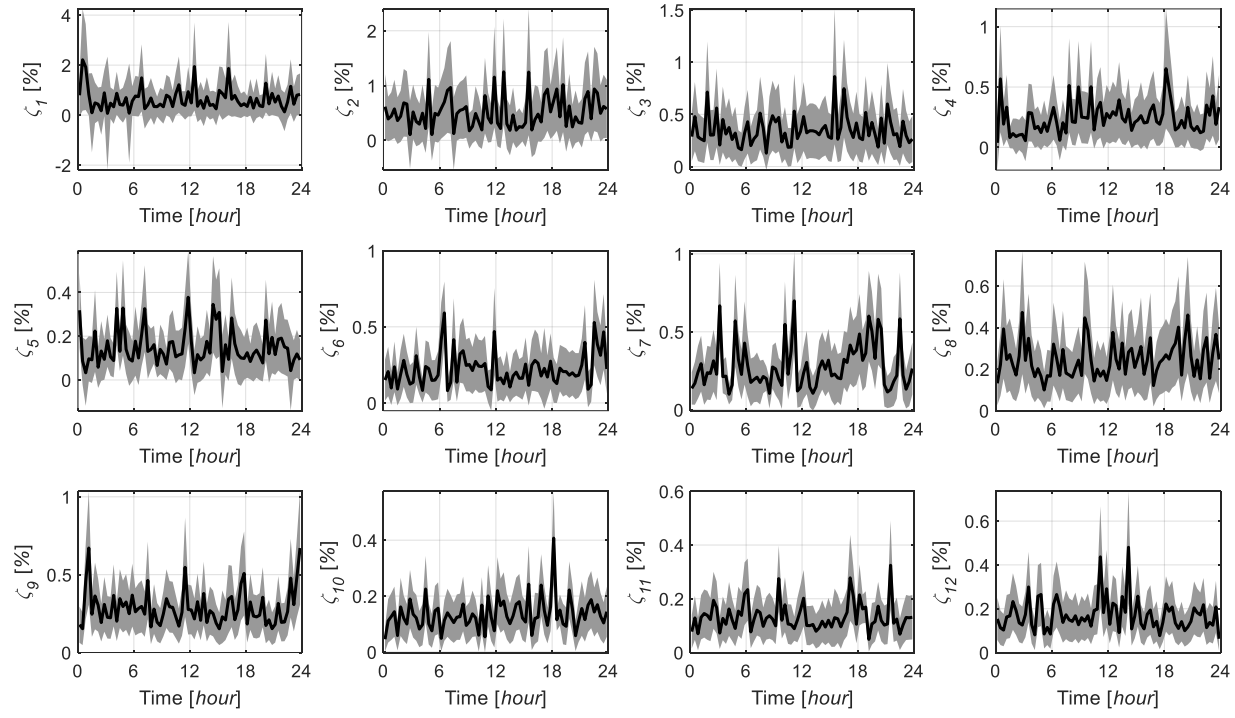


Figure 9: Identified modal shapes, Guangzhou New TV Tower.



a) Frequency



b) Damping ratio

Figure 10: Identified modal parameters for different time, Guangzhou New TV Tower.
Solid line: MPV, shaded area: ±two standard deviations.

Table 4: Time variability of modal parameters, Guangzhou New TV Tower.

Mode		1	2	3	4	5	6	7	8	9	10	11	12
Frequency c.o.v.	Sample	4.4	2.7	1.9	1.2	0.90	1.2	1.0	1.4	1.4	0.61	0.48	1.1
[‰]	Posterior	3.1	2.3	1.1	0.91	0.63	0.78	0.85	0.71	0.68	0.41	0.39	0.45
Damping ratio	Sample	67	55	39	51	54	48	51	34	39	41	38	42
c.o.v. [%]	Posterior	57	52	39	46	45	39	33	33	29	32	32	31

Note: posterior c.o.v. represnts the median of c.o.v.s of all 72 data sets.

6 Conclusions

This paper has presented an EM algorithm for the MPV computation in the Bayesian FFT method. Regarding the modal response as a latent variable, the Bayesian FFT model is formulated as a latent variable model, where the EM algorithm can be applied to maximize the log-posterior. In order to free up the norm constraints on mode shapes, the PX-EM is introduced, allowing the closed-form analytical update of all parameters except the frequency and damping ratio. The P-EM algorithm is further adopted to accelerate the convergence when there are two or more modes in the selected band. Further robustness for accurate computation is introduced based on the QR and Cholesky decompositions. The proposed EM algorithm has been verified and applied with synthetic, laboratory and field test data.

The EM algorithm gives almost identical results as the direct method for all the examples considered. Though more iterations are needed, it takes less computation time to achieve the same MPV, because, except for frequencies and damping ratios, all modal parameters can be updated analytically. For the same reason, the derivation is more systematic and elegant than the direct method. The algorithm offers a more efficient and robust (in terms of convergence) alternative that can be especially useful when the direct method has difficulty to converge. It should be mentioned that the EM algorithm does not provide a convenient way to compute the posterior covariance matrix, which can be seen as a drawback of the proposed approach. Application of the EM algorithm to other cases, e.g., multi-modes multi-setups and complex modes, is being explored.

7 Acknowledgments

Most of the work in this paper was performed when the first author was a research associate and the second author was a professor at the University of Liverpool. The financial support from the UK Engineering and Physical Sciences Research Council (EP/N017897/1) is gratefully acknowledged. The start-up fund from Zhejiang University (130000-171207704/018) is also acknowledged.

8 Appendix A. Useful results in matrix theory

This appendix summarizes the results in linear algebra that are useful in the development of the paper. A good reference can be found in [51].

1. Matrix inversion lemma

For any complex matrices $\boldsymbol{A}$, $\boldsymbol{C}$, $\boldsymbol{U}$, and $\boldsymbol{V}$ of appropriate size, with $\boldsymbol{A}$ and $\boldsymbol{C}$ invertible,

$$(\boldsymbol{A}+\boldsymbol{UCV})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1} \tag{A. 1}$$

$$(\boldsymbol{A}+\boldsymbol{UCV})^{-1}\boldsymbol{UC}=\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1} \tag{A. 2}$$

2. Matrix determinant theorem

For any complex matrices $\boldsymbol{A}$, $\boldsymbol{C}$, $\boldsymbol{U}$, and $\boldsymbol{V}$ of appropriate size, with $\boldsymbol{A}$ and $\boldsymbol{C}$ invertible,

$$|\boldsymbol{A}+\boldsymbol{UCV}|=|\boldsymbol{A}||\boldsymbol{C}||\boldsymbol{C}^{-1}+\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U}| \tag{A. 3}$$

3. Derivatives of trace

For any complex matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ of appropriate size,

$$\frac{\partial}{\partial\mathrm{Re}(\boldsymbol{A})}\mathrm{tr}(\boldsymbol{AB})=\boldsymbol{B}^{\mathrm{T}},\ \frac{\partial}{\partial\mathrm{Im}(\boldsymbol{A})}\mathrm{tr}(\boldsymbol{AB})=\mathbf{i}\boldsymbol{B}^{\mathrm{T}} \tag{A. 4}$$

$$\frac{\partial}{\partial\mathrm{Re}(\boldsymbol{A})}\mathrm{tr}(\boldsymbol{ABA}^{*}\boldsymbol{C})=\boldsymbol{CAB}+\boldsymbol{C}^{\mathrm{T}}\bar{\boldsymbol{A}}\boldsymbol{B}^{\mathrm{T}},\ \frac{\partial}{\partial\mathrm{Im}(\boldsymbol{A})}\mathrm{tr}(\boldsymbol{ABA}^{*}\boldsymbol{C})=-\mathbf{i}(\boldsymbol{CAB}-\boldsymbol{C}^{\mathrm{T}}\bar{\boldsymbol{A}}\boldsymbol{B}^{\mathrm{T}}) \tag{A. 5}$$

When $\boldsymbol{A}$ and $\boldsymbol{B}$ are both Hermitian

$$\frac{\partial}{\partial\mathrm{Re}(\boldsymbol{A})}\mathrm{tr}(\boldsymbol{AB})=2\mathrm{Re}\boldsymbol{B}-\mathrm{Re}\boldsymbol{B}\circ\boldsymbol{I},\ \frac{\partial}{\partial\mathrm{Im}(\boldsymbol{A})}\mathrm{tr}(\boldsymbol{AB})=2\mathrm{Im}\boldsymbol{B} \tag{A. 6}$$

where '$\circ$' represents matrix elements product.

3. Derivatives of log determinant

For any complex matrix $\boldsymbol{A}$,

$$\frac{\partial}{\partial\mathrm{Re}(\boldsymbol{A})}\log|\boldsymbol{A}|=\boldsymbol{A}^{-\mathrm{T}},\ \frac{\partial}{\partial\mathrm{Im}(\boldsymbol{A})}\log|\boldsymbol{A}|=\mathbf{i}\boldsymbol{A}^{-\mathrm{T}} \tag{A. 7}$$

Especially, when $\boldsymbol{A}$ is Hermitian,

$$\frac{\partial}{\partial\mathrm{Re}(\boldsymbol{A})}\log|\boldsymbol{A}|=2\mathrm{Re}\boldsymbol{A}^{-1}-\mathrm{Re}\boldsymbol{A}^{-1}\circ\boldsymbol{I},\ \frac{\partial}{\partial\mathrm{Im}(\boldsymbol{A})}\log|\boldsymbol{A}|=2\mathrm{Im}\boldsymbol{A}^{-1} \tag{A. 8}$$

9 Appendix B. Proof of equations

In this appendix, we will provide proofs for various equations which appear in Section 0.

1. Proof of Eqn. (15)

Recall $\widehat{\boldsymbol{\mathcal{F}}}_k=\boldsymbol{\Phi}\boldsymbol{\eta}_k+\boldsymbol{\varepsilon}_k$. Given $\boldsymbol{\theta}$ and $\boldsymbol{\eta}_k$, the first term $\boldsymbol{\Phi}\boldsymbol{\eta}_k$ is deterministic and so the uncertainty of $\widehat{\boldsymbol{\mathcal{F}}}_k$ comes from the noise term $\boldsymbol{\varepsilon}_k$. Since the modal force and noise are assumed to be independent, knowing $\boldsymbol{\eta}_k$ does not provide additional information for $\boldsymbol{\varepsilon}_k$ beyond that by $\boldsymbol{\theta}$. Consequently, $\boldsymbol{\varepsilon}_k$ is still complex Gaussian with zero mean and covariance matrix $S_e\boldsymbol{I}_n$, and so

$$\mathrm{E}_{\widehat{\boldsymbol{\mathcal{F}}}_k|\boldsymbol{\eta}_k,\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\mathcal{F}}}_k\right]=\boldsymbol{\Phi}\boldsymbol{\eta}_k \tag{B. 1}$$

$$\mathrm{Cov}_{\widehat{\boldsymbol{\mathcal{F}}}_k|\boldsymbol{\eta}_k,\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\mathcal{F}}}_k\right]=S_e\boldsymbol{I}_n \tag{B. 2}$$

Knowing $\boldsymbol{\eta}_k|\boldsymbol{\theta}\sim\mathcal{CN}(\mathbf{0},\boldsymbol{H}_k)$, we can compute the complete-data log-likelihood

$$L\left(\boldsymbol{\theta}\middle|\left\{\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\eta}_k\right\}\right)=\log p\left(\left\{\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\eta}_k\right\}\middle|\boldsymbol{\theta}\right)=\sum_{k=1}^{N_f}\log p\left(\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\eta}_k\middle|\boldsymbol{\theta}\right) \tag{B. 3}$$

$$= \sum_{k=1}^{N_f} \log\left[p\left(\widehat{\boldsymbol{\mathcal{F}}}_k|\boldsymbol{\eta}_k,\boldsymbol{\theta}\right)p(\boldsymbol{\eta}_k|\boldsymbol{\theta})\right] = \sum_{k=1}^{N_f} \log p\left(\widehat{\boldsymbol{\mathcal{F}}}_k|\boldsymbol{\eta}_k,\boldsymbol{\theta}\right) + \sum_{k=1}^{N_f} \log p(\boldsymbol{\eta}_k|\boldsymbol{\theta})$$
$$= -2nN_f \log \pi - nN_f \log S_e - S_e^{-1} \sum_{k=1}^{N_f} \left[\widehat{\boldsymbol{\mathcal{F}}}_k - \boldsymbol{\Phi}\boldsymbol{\eta}_k\right]^* \left[\widehat{\boldsymbol{\mathcal{F}}}_k - \boldsymbol{\Phi}\boldsymbol{\eta}_k\right] + \sum_{k=1}^{N_f} \log|\boldsymbol{H}_k^{-1}| - \sum_{k=1}^{N_f} \boldsymbol{\eta}_k^* \boldsymbol{H}_k^{-1} \boldsymbol{\eta}_k$$

2. Proof of Eqns. (16) and (17)

Given $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\mathcal{F}}}_k$, it is clear that the distribution of $\boldsymbol{\eta}_k$ is still complex Gaussian. Its conditional mean and covariance matrix can be obtained by using standard results in multivariate complex Gaussian distribution. In particular, let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be two zero mean jointly complex Gaussian vectors. Given $\boldsymbol{X}_1$, the conditional mean and covariance matrix of $\boldsymbol{X}_2$ are respectively $\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{X}_1$ and $\boldsymbol{C}_{22} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}$, where $\boldsymbol{C}_{ij}$ denotes the (i,j)-partition of the covariance matrix of $\{\boldsymbol{X}_1, \boldsymbol{X}_2\}$. Applying this result with $\boldsymbol{X}_1 = \widehat{\boldsymbol{\mathcal{F}}}_k$ and $\boldsymbol{X}_2 = \boldsymbol{\eta}_k$, and using their covariance matrix in Eqn. (14), the conditional mean and covariance matrix of $\boldsymbol{\eta}_k$ are given by

$$\begin{aligned} &\mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}}[\boldsymbol{\eta}_k] = \boldsymbol{H}_k\boldsymbol{\Phi}^{\mathrm{T}}[\boldsymbol{\Phi}\boldsymbol{H}_k\boldsymbol{\Phi}^{\mathrm{T}} + S_e\boldsymbol{I}_n]^{-1}\widehat{\boldsymbol{\mathcal{F}}}_k \\ &= S_e^{-1}(\boldsymbol{H}_k^{-1} + S_e^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\widehat{\boldsymbol{\mathcal{F}}}_k = (S_e\boldsymbol{H}_k^{-1} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\widehat{\boldsymbol{\mathcal{F}}}_k \end{aligned} \tag{B. 4}$$

where we have used the property of Eqn. (A. 2) in deriving the second equation;

$$\begin{aligned} &\mathrm{Cov}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}}[\boldsymbol{\eta}_k] = \boldsymbol{H}_k - \boldsymbol{H}_k\boldsymbol{\Phi}^{\mathrm{T}}[\boldsymbol{\Phi}\boldsymbol{H}_k\boldsymbol{\Phi}^{\mathrm{T}} + S_e\boldsymbol{I}_n]^{-1}\boldsymbol{\Phi}\boldsymbol{H}_k \\ &= (\boldsymbol{H}_k^{-1} + S_e^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1} = S_e(S_e\boldsymbol{H}_k^{-1} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1} \end{aligned} \tag{B. 5}$$

where the matrix inversion lemma Eqn. (A. 1) has been applied in deriving the second equation.

3. Proof of Eqn. (20)

Similar to the proof of Eqn. (15), we have $\widehat{\boldsymbol{\mathcal{F}}}_k|(\boldsymbol{\eta}_k,\boldsymbol{\theta}_*,\boldsymbol{\gamma}) \sim \mathcal{CN}(\boldsymbol{\Phi}_*\boldsymbol{\gamma}\boldsymbol{\eta}_k, S_{e*}\boldsymbol{I}_n)$ and $\boldsymbol{\eta}_k|\boldsymbol{\theta}_* \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{H}_k)$ for the augmented model. The complete-data log-likelihood is then given as

$$\begin{aligned} &L\left(\boldsymbol{\theta}_*,\boldsymbol{\gamma}|\{\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\eta}_k\}\right) = \log p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\eta}_k\}|\boldsymbol{\theta}_*,\boldsymbol{\gamma}\right) \\ &= \sum_{k=1}^{N_f} \log p\left(\widehat{\boldsymbol{\mathcal{F}}}_k|\boldsymbol{\eta}_k,\boldsymbol{\theta}_*,\boldsymbol{\gamma}\right) + \sum_{k=1}^{N_f} \log p(\boldsymbol{\eta}_k|\boldsymbol{\theta}_*) \\ &= -2nN_f \log \pi - nN_f \log S_{e*} - S_{e*}^{-1} \sum_{k=1}^{N_f} \left[\widehat{\boldsymbol{\mathcal{F}}}_k - \boldsymbol{\Phi}_*\boldsymbol{\gamma}\boldsymbol{\eta}_k\right]^* \left[\widehat{\boldsymbol{\mathcal{F}}}_k - \boldsymbol{\Phi}_*\boldsymbol{\gamma}\boldsymbol{\eta}_k\right] + \\ &\sum_{k=1}^{N_f} \log|\boldsymbol{H}_k^{-1}| - \sum_{k=1}^{N_f} \boldsymbol{\eta}_k^* \boldsymbol{H}_k^{-1} \boldsymbol{\eta}_k \end{aligned} \tag{B. 6}$$

Taking the expectation of the above equation w.r.t. the distribution of $\boldsymbol{\eta}_k$ conditioning on $\widehat{\boldsymbol{\mathcal{F}}}_k$, $\boldsymbol{\theta}^{(t)}$, and $\boldsymbol{\gamma}_0$, one can evaluate the Q-function as

$$\begin{aligned} &Q\left(\boldsymbol{\theta}_*,\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0\right) = \mathrm{E}_{\{\boldsymbol{\eta}_k\}|\{\widehat{\boldsymbol{\mathcal{F}}}_k\},\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}\left[L\left(\boldsymbol{\theta}_*,\boldsymbol{\gamma}|\{\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\eta}_k\}\right)\right] \\ &= -2nN_f \log \pi - nN_f \log S_{e*} - S_{e*}^{-1} \sum_{k=1}^{N_f} \mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}\left\{\left[\widehat{\boldsymbol{\mathcal{F}}}_k - \boldsymbol{\Phi}_*\boldsymbol{\gamma}\boldsymbol{\eta}_k\right]^* \left[\widehat{\boldsymbol{\mathcal{F}}}_k - \right.\right. \\ &\left.\left.\boldsymbol{\Phi}_*\boldsymbol{\gamma}\boldsymbol{\eta}_k\right]\right\} + \sum_{k=1}^{N_f} \log|\boldsymbol{H}_k^{-1}| - \sum_{k=1}^{N_f} \mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}[\boldsymbol{\eta}_k^* \boldsymbol{H}_k^{-1} \boldsymbol{\eta}_k] \\ &= -2nN_f \log \pi - nN_f \log S_{e*} - S_{e*}^{-1} \sum_{k=1}^{N_f} \mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}\left\{\widehat{\boldsymbol{\mathcal{F}}}_k^* \widehat{\boldsymbol{\mathcal{F}}}_k - \widehat{\boldsymbol{\mathcal{F}}}_k^* \boldsymbol{\Phi}_*\boldsymbol{\gamma}\boldsymbol{\eta}_k - \right. \\ &\left.\boldsymbol{\eta}_k^* \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Phi}_*^{\mathrm{T}} \widehat{\boldsymbol{\mathcal{F}}}_k + \boldsymbol{\eta}_k^* \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Phi}_*^{\mathrm{T}} \boldsymbol{\Phi}_*\boldsymbol{\gamma}\boldsymbol{\eta}_k\right\} + \sum_{k=1}^{N_f} \log|\boldsymbol{H}_k^{-1}| - \sum_{k=1}^{N_f} \mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0}[\mathrm{tr}(\boldsymbol{\eta}_k\boldsymbol{\eta}_k^* \boldsymbol{H}_k^{-1})] \end{aligned} \tag{B. 7}$$

$$= -2nN_f \log \pi - nN_f \log S_{e*} - S_{e*}^{-1} Q_1\left(\boldsymbol{\Phi}_*, \boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right) + Q_2\left(\boldsymbol{f}_*, \boldsymbol{\zeta}_*, \boldsymbol{S}_* | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right)$$

where

$$Q_1\left(\boldsymbol{\Phi}_*, \boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right) = \sum_{k=1}^{N_f} \widehat{\boldsymbol{\mathcal{F}}}_k^* \widehat{\boldsymbol{\mathcal{F}}}_k - 2\mathrm{tr}\left[\boldsymbol{\Phi}_* \boldsymbol{\gamma} \sum_{k=1}^{N_f} \mathrm{Re}\left(\boldsymbol{w}_{1k} \widehat{\boldsymbol{\mathcal{F}}}_k^*\right)\right] + \mathrm{tr}\left[\boldsymbol{\Phi}_* \boldsymbol{\gamma} \mathrm{Re}\left(\sum_{k=1}^{N_f} \boldsymbol{w}_{2k}\right) \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Phi}_*^{\mathrm{T}}\right] \quad \text{(B. 8)}$$

$$Q_2\left(\boldsymbol{f}_*, \boldsymbol{\zeta}_*, \boldsymbol{S}_* | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right) = \sum_{k=1}^{N_f} \log|\boldsymbol{H}_k^{-1}| - \mathrm{tr}\left[\sum_{k=1}^{N_f} \boldsymbol{H}_k^{-1} \boldsymbol{w}_{2k}\right] \quad \text{(B. 9)}$$

$$\boldsymbol{w}_{1k} = \mathrm{E}_{\boldsymbol{\eta}_k | \widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0}[\boldsymbol{\eta}_k], \boldsymbol{w}_{2k} = \mathrm{E}_{\boldsymbol{\eta}_k | \widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0}[\boldsymbol{\eta}_k \boldsymbol{\eta}_k^*] \quad \text{(B. 10)}$$

4. Proof of Eqn. (25)

To obtain the optimal value of S_{e*}, taking the partial derivative of $Q\left(\boldsymbol{\theta}_*, \boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right)$ in Eqn. (20) w.r.t. S_{e*}^{-1} gives

$$\frac{\partial Q\left(\boldsymbol{\theta}_*, \boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right)}{\partial S_{e*}^{-1}} = nN_f S_{e*} - Q_1\left(\boldsymbol{\Phi}_*, \boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right) \quad \text{(B. 11)}$$

Setting it to zero and solving for S_{e*} yields

$$S_{e*} = \frac{1}{nN_f} Q_1\left(\boldsymbol{\Phi}_*, \boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right) \quad \text{(B. 12)}$$

Substituting Eqn. (24) into Eqn. (21) gives

$$\begin{aligned} &Q_1\left(\boldsymbol{\Phi}_*, \boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right) = \mathrm{tr}\left[\sum_{k=1}^{N_f} \mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k \widehat{\boldsymbol{\mathcal{F}}}_k^*\right)\right] - \\ &2\mathrm{tr}\left[\sum_{k=1}^{N_f} \mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k \boldsymbol{w}_{1k}^*\right) \left[\sum_{k=1}^{N_f} \mathrm{Re}(\boldsymbol{w}_{2k})\right]^{-1} \sum_{k=1}^{N_f} \mathrm{Re}\left(\boldsymbol{w}_{1k} \widehat{\boldsymbol{\mathcal{F}}}_k^*\right)\right] + \\ &\mathrm{tr}\left[\sum_{k=1}^{N_f} \mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k \boldsymbol{w}_{1k}^*\right) \left[\sum_{k=1}^{N_f} \mathrm{Re}(\boldsymbol{w}_{2k})\right]^{-1} \sum_{k=1}^{N_f} \mathrm{Re}\left(\boldsymbol{w}_{1k} \widehat{\boldsymbol{\mathcal{F}}}_k^*\right)\right] \end{aligned} \quad \text{(B. 13)}$$

$$= \mathrm{tr}\left[\sum_{k=1}^{N_f} \mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k \widehat{\boldsymbol{\mathcal{F}}}_k^*\right)\right] - \mathrm{tr}\left[\sum_{k=1}^{N_f} \mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k \boldsymbol{w}_{1k}^*\right) \left[\sum_{k=1}^{N_f} \mathrm{Re}(\boldsymbol{w}_{2k})\right]^{-1} \sum_{k=1}^{N_f} \mathrm{Re}\left(\boldsymbol{w}_{1k} \widehat{\boldsymbol{\mathcal{F}}}_k^*\right)\right]$$

$$= \mathrm{tr}\left[\sum_{k=1}^{N_f} \mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k \widehat{\boldsymbol{\mathcal{F}}}_k^*\right) - \boldsymbol{\Phi}' \sum_{k=1}^{N_f} \mathrm{Re}(\boldsymbol{w}_{2k}) \boldsymbol{\Phi}'^{\mathrm{T}}\right]$$

Therefore,

$$S_{e*} = \frac{1}{nN_f} \mathrm{tr}\left[\sum_{k=1}^{N_f} \mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k \widehat{\boldsymbol{\mathcal{F}}}_k^*\right) - \boldsymbol{\Phi}' \sum_{k=1}^{N_f} \mathrm{Re}(\boldsymbol{w}_{2k}) \boldsymbol{\Phi}'^{\mathrm{T}}\right] \quad \text{(B. 14)}$$

4. Proof of Eqn. (26)

Before proceeding to the derivation, we first rewrite Eqn. (22) as

$$Q_2\left(\boldsymbol{f}_*, \boldsymbol{\zeta}_*, \boldsymbol{S}_* | \boldsymbol{\theta}^{(t)}, \boldsymbol{\gamma}_0\right) = \sum_{k=1}^{N_f} \log|\boldsymbol{h}_k^{-1} \boldsymbol{h}_k^{-*}| + N_f \log|\boldsymbol{S}_*^{-1}| - \mathrm{tr}\left(\boldsymbol{S}_*^{-1} \sum_{k=1}^{N_f} \boldsymbol{h}_k^{-1} \boldsymbol{w}_{2k} \boldsymbol{h}_k^{-*}\right) \quad \text{(B. 15)}$$

Using Eqns. (A. 6) and (A. 8), we have the partial derivative

$$\frac{\partial Q(\boldsymbol{\theta}_*,\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)},\boldsymbol{\gamma}_0)}{\partial \mathrm{Re}(S_*^{-1})} = N_f[2\mathrm{Re}(\boldsymbol{S}_*) - \mathrm{Re}(\boldsymbol{S}_*) \circ \boldsymbol{I}] - \left[2\mathrm{Re}\left(\sum_{k=1}^{N_f} \boldsymbol{h}_k^{-1}\boldsymbol{w}_{2k}\boldsymbol{h}_k^{-*}\right) - \mathrm{Re}\left(\sum_{k=1}^{N_f} \boldsymbol{h}_k^{-1}\boldsymbol{w}_{2k}\boldsymbol{h}_k^{-*}\right) \circ \boldsymbol{I}\right] \quad \text{(B. 16)}$$

and setting it to zero gives

$$\mathrm{Re}(\boldsymbol{S}_*) = \frac{1}{N_f}\mathrm{Re}\left(\sum_{k=1}^{N_f} \boldsymbol{h}_k^{-1}\boldsymbol{w}_{2k}\boldsymbol{h}_k^{-*}\right) \quad \text{(B. 17)}$$

Similarly, we can derive

$$\mathrm{Im}(\boldsymbol{S}_*) = \frac{1}{N_f}\mathrm{Im}\left(\sum_{k=1}^{N_f} \boldsymbol{h}_k^{-1}\boldsymbol{w}_{2k}\boldsymbol{h}_k^{-*}\right) \quad \text{(B. 18)}$$

and combining Eqn. (B. 17) yields

$$\boldsymbol{S}_* = \frac{1}{N_f}\sum_{k=1}^{N_f} \boldsymbol{h}_k^{-1}\boldsymbol{w}_{2k}\boldsymbol{h}_k^{-*} \quad \text{(B. 19)}$$

10 References

[1] E. Reynders, System Identification Methods for (Operational) Modal Analysis: Review and Comparison, Arch. Comput. Methods Eng. 19 (2012) 51–124. doi:10.1007/s11831-012-9069-x.

[2] R. Brincker, C. Ventura, Introduction to operational modal analysis, Wiley, Chichester, 2015.

[3] S.-K. Au, Operational modal analysis : modeling, bayesian inference, uncertainty laws, Springer, Singapore, 2017.

[4] H. Katsuchi, H. Yamada, S. Kusuhara, Structural Monitoring and Design Verification of Akashi Kaikyo Bridge, in: 11th Bienn. ASCE Aerosp. Div. Int. Conf. Eng. Sci. Constr. Oper. Challenging Environ., Long Beach, USA, 2008: pp. 1–8.

[5] J. Brownjohn, K.Y. Koo, C. Basagiannis, A. Alskif, A. Ngonda, Vibration monitoring and condition assessment of the University of Sheffield Arts Tower during retrofit, J. Civ. Struct. Heal. Monit. (2013). doi:10.1007/s13349-012-0028-x.

[6] B.F.J. Spencer, S. Nagarajaiah, State of the Art of Structural Control, J. Struct. Eng. (2003). doi:10.1061/(ASCE)0733-9445(2003)129:7(845).

[7] J.M.W. Brownjohn, E.P. Carden, C.R. Goddard, G. Oudin, Real-time performance monitoring of tuned mass damper system for a 183 m reinforced concrete chimney, J. Wind Eng. Ind. Aerodyn. (2010). doi:10.1016/j.jweia.2009.10.013.

[8] E. Ntotsios, C. Papadimitriou, P. Panetsos, G. Karaiskos, K. Perros, P.C. Perdikaris, Bridge health monitoring system based on vibration measurements, Bull. Earthq. Eng. 7 (2008) 469–483. doi:10.1007/s10518-008-9067-4.

[9] J.M.W. Brownjohn, Structural health monitoring of civil infrastructure, Philos. Trans. R.

Soc. London A Math. Phys. Eng. Sci. 365 (2007) 589–622. doi:10.1098/rsta.2006.1925.

[10] A. Cole, On-line failure detection and damping measurement of aerospace structures by random decrement signatures, Mountain View, USA, 1973. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19730010202.pdf.

[11] J.S. Bendat, A.G. Piersol, Engineering Applications of Correlation and Spectral Analysis, 1st ed., John Wiley & Sons, Ltd, New York, USA, 1980.

[12] P. Van Overschee, B. De Moor, Subspace Identification for Linear Systems, Springer US, Boston, USA, 1996. doi:10.1007/978-1-4613-0465-4.

[13] R. Brincker, L. Zhang, P. Andersen, Modal identification from ambient responses using frequency domain decomposition, in: 18th Int. Modal Anal. Conf., San Antonio, USA, 2000: pp. 625–630.

[14] E. Reynders, R. Pintelon, G. De Roeck, Uncertainty bounds on modal parameters obtained from stochastic subspace identification, Mech. Syst. Signal Process. 22 (2008) 948–969. doi:10.1016/j.ymssp.2007.10.009.

[15] M. Döhler, X.-B. Lam, L. Mevel, Uncertainty quantification for modal parameters from stochastic subspace identification on multi-setup measurements, Mech. Syst. Signal Process. 36 (2013) 562–581. doi:10.1016/J.YMSSP.2012.11.011.

[16] E. Reynders, K. Maes, G. Lombaert, G. De Roeck, Uncertainty quantification in operational modal analysis with stochastic subspace identification: Validation and applications, Mech. Syst. Signal Process. 66–67 (2016) 13–30. doi:10.1016/J.YMSSP.2015.04.018.

[17] S.-K. Au, F.-L. Zhang, Y.-C. Ni, Bayesian operational modal analysis: Theory, computation, practice, Comput. Struct. 126 (2013) 3–14. doi:10.1016/j.compstruc.2012.12.015.

[18] K.-V. Yuen, L.S. Katafygiotis, Bayesian time–domain approach for modal updating using ambient data, Probabilistic Eng. Mech. 16 (2001) 219–231. doi:10.1016/S0266-8920(01)00004-2.

[19] B. Li, A. Der Kiureghian, Operational modal identification using variational Bayes, Mech. Syst. Signal Process. 88 (2017) 377–398. doi:10.1016/j.ymssp.2016.11.007.

[20] B. Li, A. Der Kiureghian, S.-K. Au, A Gibbs sampling algorithm for structural modal identification under seismic excitation, Earthq. Eng. Struct. Dyn. (2018) 1–21. doi:10.1002/eqe.3094.

[21] L.S. Katafygiotis, K.-V. Yuen, Bayesian spectral density approach for modal updating using ambient data, Earthq. Eng. Struct. Dyn. 30 (2001) 1103–1123. doi:10.1002/eqe.53.

[22] K.-V. Yuen, L. Katafygiotis, Bayesian fast Fourier transform approach for modal updating using ambient data, Adv. Struct. Eng. 6 (2003) 81–95. doi:10.1260/136943303769013183.

[23] S.-K. Au, Fast Bayesian FFT Method for Ambient Modal Identification with Separated Modes, J. Eng. Mech. 137 (2011) 214–226. doi:10.1061/(ASCE)EM.1943-7889.0000213.

[24] F.-L. Zhang, S.-K. Au, Erratum for “Fast Bayesian FFT Method for Ambient Modal Identification with Separated Modes” by Siu-Kui Au, J. Eng. Mech. 139 (2013) 545–545.

doi:10.1061/(ASCE)EM.1943-7889.0000501.

[25] S.-K. Au, Fast Bayesian ambient modal identification in the frequency domain, Part I: Posterior most probable value, Mech. Syst. Signal Process. 26 (2012) 60–75. doi:10.1016/j.ymssp.2011.06.017.

[26] S.-K. Au, Fast Bayesian ambient modal identification in the frequency domain, Part II: Posterior uncertainty, Mech. Syst. Signal Process. 26 (2012) 76–90. doi:10.1016/j.ymssp.2011.06.019.

[27] W.-J. Yan, L.S. Katafygiotis, A two-stage fast Bayesian spectral density approach for ambient modal analysis. Part II: Mode shape assembly and case studies, Mech. Syst. Signal Process. 54–55 (2015) 156–171. doi:10.1016/J.YMSSP.2014.08.016.

[28] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, J. R. Stat. Soc. Ser. B. 39 (1977) 1–38. doi:10.2307/2984875.

[29] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, 2nd ed., Wiley, New York, USA, 2008.

[30] F.J. Cara, J. Carpio, J. Juan, E. Alarcón, An approach to operational modal analysis using the expectation maximization algorithm, Mech. Syst. Signal Process. 31 (2012) 109–129. doi:10.1016/j.ymssp.2012.04.004.

[31] T.J. Matarazzo, S.N. Pakzad, STRIDE for Structural Identification Using Expectation Maximization: Iterative Output-Only Method for Modal Identification, J. Eng. Mech. 142 (2016) 04015109. doi:10.1061/(ASCE)EM.1943-7889.0000951.

[32] B. Li, A. Der Kiureghian, Latent variable model and its application in operational modal analysis, in: 12th Int. Conf. Struct. Saf. Reliab., Vienna, Austria, 2017: pp. 2748–2757.

[33] C.F.J. Wu, On the Convergence Properties of the EM Algorithm, Ann. Stat. (1983). doi:10.1214/aos/1176346060.

[34] C. Liu, D.B. Rubin, Y.N. Wu, Parameter Expansion to Accelerate EM: The PX-EM Algorithm, Biometrika. 85 (1998) 755–770. doi:10.1093/biomet/85.4.755.

[35] A. Berlinet, C. Roland, Parabolic acceleration of the EM algorithm, Stat. Comput. (2009). doi:10.1007/s11222-008-9067-x.

[36] S.-K. Au, Model validity and frequency band selection in operational modal analysis, Mech. Syst. Signal Process. 81 (2016) 339–359. doi:10.1016/j.ymssp.2016.03.025.

[37] D.R. Brillinger, Time series : data analysis and theory, SIAM, Philadelphia, 2001. doi:http://dx.doi.org/10.1137/1.9780898719246.

[38] J.L. Beck, Bayesian system identification based on probability logic, Struct. Control Heal. Monit. 17 (2010) 825–847. doi:10.1002/stc.424.

[39] R.W. Keener, Theoretical statistics : topics for a core course, Springer, New York, 2010.

[40] H.H. Andersen, M. Højbjerre, D. Sørensen, P.S. Eriksen, The Multivariate Complex Normal Distribution, in: Springer, New York, NY, 1995: pp. 15–37. doi:10.1007/978-1-

4612-4240-6_2.

[41] MATLAB and Optimization Toolbox Release 2016b, The MathWorks, Inc., Natick, Massachusetts, United States, 2016.

[42] J.C. Lagarias, J.A. Reeds, M.H. Wright, P.E. Wright, Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions, SIAM J. Optim. (1998). doi:10.1137/S1052623496303470.

[43] S.-K. Au, F.-L. Zhang, On assessing the posterior mode shape uncertainty in ambient modal identification, Probabilistic Eng. Mech. 26 (2011) 427–434. doi:10.1016/j.probengmech.2010.11.009.

[44] Y.Q. Ni, Y. Xia, W.Y. Liao, J.M. Ko, Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower, Struct. Control Heal. Monit. (2009). doi:10.1002/stc.303.

[45] Y.Q. Ni, Y. Xia, W. Lin, W.H. Chen, J.M. Ko, SHM benchmark for high-rise structures: A reduced-order finite element model and field measurement data, Smart Struct. Syst. (2012). doi:10.12989/sss.2012.10.4_5.411.

[46] W.H. Chen, Z.R. Lu, W. Lin, S.H. Chen, Y.Q. Ni, Y. Xia, W.Y. Liao, Theoretical and experimental modal analysis of the Guangzhou New TV Tower, Eng. Struct. (2011). doi:10.1016/j.engstruct.2011.07.028.

[47] X. Ye, Q. Yan, W. Wang, X. Yu, Modal identification of Canton Tower under uncertain environmental conditions, Smart Struct. Syst. (2012). doi:10.12989/sss.2012.10.4_5.353.

[48] S.-C. Kuok, K.-V. Yuen, Structural health monitoring of Canton Tower using Bayesian framework, Smart Struct. Syst. (2012). doi:10.12989/sss.2012.10.4_5.375.

[49] F.L. Zhang, Y.Q. Ni, Y.C. Ni, Y.W. Wang, Operational modal analysis of Canton Tower by a fast frequency domain Bayesian method, Smart Struct. Syst. (2016). doi:10.12989/sss.2016.17.2.209.

[50] S.-K. Au, B. Li, Posterior uncertainty, asymptotic law and Cramér-Rao bound, Struct. Control Heal. Monit. 25 (2018) e2113. doi:10.1002/stc.2113.

[51] S.R. Searle, Matrix Algebra Useful for Statistics, Technometrics. 27 (1982) 438. doi:10.2307/1270478.